SANDIA REPORT

SAND2004-5082
Unlimited Release
Printed October 2004

Methods for model selection in applied science and engineering

Richard V. Field, Jr.

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of Energy's
National Nuclear Security Administration under Contract DE-AC04-94-AL85000.

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.doe.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online ordering: http://www.ntis.gov/ordering.htm

SAND2004-5082
Unlimited Release
Printed October 2004

Methods for model selection in applied science and engineering

Richard V. Field, Jr.
Structural Dynamics Research Department
Sandia National Laboratories
P.O. Box 5800
Albuquerque, NM 87185-0847
rvfield@sandia.gov

Abstract

Mathematical models are developed and used to study the properties of complex systems and/or modify these systems to satisfy some performance requirements in just about every area of applied science and engineering. A particular reason for developing a model, *e.g.*, performance assessment or design, is referred to as the model use. Our objective is the development of a methodology for selecting a model that is sufficiently accurate for an intended use.

Information on the system being modeled is, in general, incomplete, so that there may be two or more models consistent with the available information. The collection of these models is called the class of candidate models. Methods are developed for selecting the optimal member from a class of candidate models for the system. The optimal model depends on the available information, the selected class of candidate models, and the model use.

Classical methods for model selection, including the method of maximum likelihood and Bayesian methods, as well as a method employing a decision-theoretic approach, are formulated to select the optimal model for numerous applications. There is no requirement that the candidate models be random. Classical methods for model selection ignore model use and require data to be available. Examples are used to show that these methods can be unreliable when data is limited. The decision-theoretic approach to model selection does not have these limitations, and model use is included through an appropriate utility function. This is especially important when modeling high risk systems, where the consequences of using an inappropriate model for the system can be disastrous.

The decision-theoretic method for model selection is developed and applied for a series of complex and diverse applications. These include the selection of the: (1) optimal order of the polynomial chaos approximation for non-Gaussian random variables and stationary stochastic processes, (2) optimal pressure load model to be applied to a spacecraft during atmospheric re-entry, and (3) optimal design of a distributed sensor network for the purpose of vehicle tracking and identification.

Acknowledgment

This dissertation could not have been completed without the help and support of many individuals. I would first like to thank Professor Mircea Grigoriu for being an excellent academic advisor, patient teacher, and good friend. I learned a great deal from him on the importance of organization, consistency, and rigor, and he provided consistent encouragement and support throughout my studies at Cornell. I would also like to thank my special committee members, Professors Katerina Papoulia and Gennady Samorodnitsky, for their technical advice and comments on my work. Thanks goes to Çağdaş Kafalı, fellow graduate student and officemate, for reading parts of my dissertation, suffering through Measure Theory with me, and entertaining me with tales of his adventures.

I wish to express my appreciation to Dr. Steve Wojtkiewicz, Dr. John Red-Horse, Dr. Garth Reese, and Dr. Sharon Deland, colleagues of mine from Sandia National Laboratories, for the useful technical discussions we've had and their encouragement and support for my decision to come back to school. I want to acknowledge the help of Dr. David Martinez, Dr. Hal Morgan, and Dr. Jim Redmond, also from Sandia. Thanks to Professor Larry Bergman of the University of Illinois and Professor Erik Johnson of the University of Southern California for their insightful comments on my research.

I gratefully acknowledge the support of Sandia's Doctoral Study Program (DSP), which provided tuition and partial salary for three years.

This page intentionally left blank.

Contents

Appendix

List of Figures

List of Tables

Preface

This report is a reproduction of the author's Ph.D. dissertation, originally published by Cornell University. The work was completed as part of Sandia's Doctoral Studies Program (DSP) during August, 2001 - July, 2004.

Chapter 1

Introduction

Essentials of the model selection problem are presented. The mathematical modeling process is discussed in Section 1.1. The construction of one or more mathematical models, where each model is consistent with all available information on the system being modeled, is considered in Section 1.2. Methods for model selection are outlined in Section 1.3.

1.1 Mathematical modeling

Consider an object or collection of objects, herein called a system, whose properties we wish to study. If the system already exists, we may be able to perform experiments to study the system properties. If the system does not exist, or cost and/or safety concerns prohibit experimentation, we need another method of study. A mathematical model for a system, defined to be a collection of mathematical equations approximating the system properties, can be used in this case.

A mathematical model can be developed based on any information we have regarding the behavior of the system; available information on the system is, in general, of two types: (1) data, and (2) prior knowledge. The former consists of a finite number of observations (measurements) of the system and/or experiments on the system. The latter is knowledge, other than data, that we have about the system. It is made up of the opinions and theories of experts, as well as any literature on the subject. Most often, the prior knowledge is expressed as a set of mathematical relations; qualitative knowledge may also be used. To illustrate these concepts, consider a model for the structural properties of aluminum. Any experimental measurements of the stiffness of the material is one example of data. Further, prior knowledge tells us that: (1) a linear, elastic model form may be appropriate, and (2) stiffness is nonnegative; the mathematical model for aluminum must reflect this.

The available information is, in general, incomplete; the amount of data is finite, and the prior knowledge may be underdeveloped and/or deficient. Even in the case of complete information, where we would understand precisely the behavior of the system, it is likely that the behavior may be too complicated to describe exactly, and simplified mathematical models for the system, *e.g.*, response surface approximations, must be used.

A mathematical model is developed for one or more reasons, what we refer to as the model use. There are many examples of model use, a few are (from [61], p. 8): description, insight, optimization, prediction, and design. For example, consider the case of a mathematical model for a cantilever beam made from the linear, elastic material discussed above. We may use this model to select the optimal beam cross section (optimization), to assess the natural frequencies of the beam (prediction), or to select the cross section so that the tip deflection due to a prescribed load is less that some critical value (design). In general, our objective is to develop a mathematical model for the system that is sufficiently accurate for the intended model use.

Kapur [61], pp. 9-13, presents some interesting characteristics of mathematical models.

- Realism: we want mathematical models to be as realistic as possible, but this may prove mathematically intractable.
- Model use: we may need different models for explaining different aspects of the same system.
- Non-uniqueness: a system need not have only one mathematical model.
- Complexity: a complex mathematical model does not necessarily provide a superior understanding of the system.
- Cost of modeling: every model can be improved, but not every improvement is worth the time and effort.

Mathematical models are developed and used to study the properties of complex systems in just about every area of applied science and engineering. The mathematical modeling procedure involves three elements: (1) the system for study, (2) the mathematical model or models for the system, and (3) the mathematical solution. Figure 1.1 illustrates the procedure, where elements (1)-(3) are enclosed by ellipses. Similar diagrams are discussed in p. 15 of [23], [69], and [87].

There are four steps to the mathematical modeling procedure; the steps are enclosed by the rectangles in Fig. 1.1. As indicated by the circular structure in Fig. 1.1, mathematical modeling is an iterative process. Construction and calculation, steps (1) and (2), are performed repeatedly until the mathematical model for the system is deemed valid by step (3). Only then may we proceed to step (4). The steps in the procedure are:

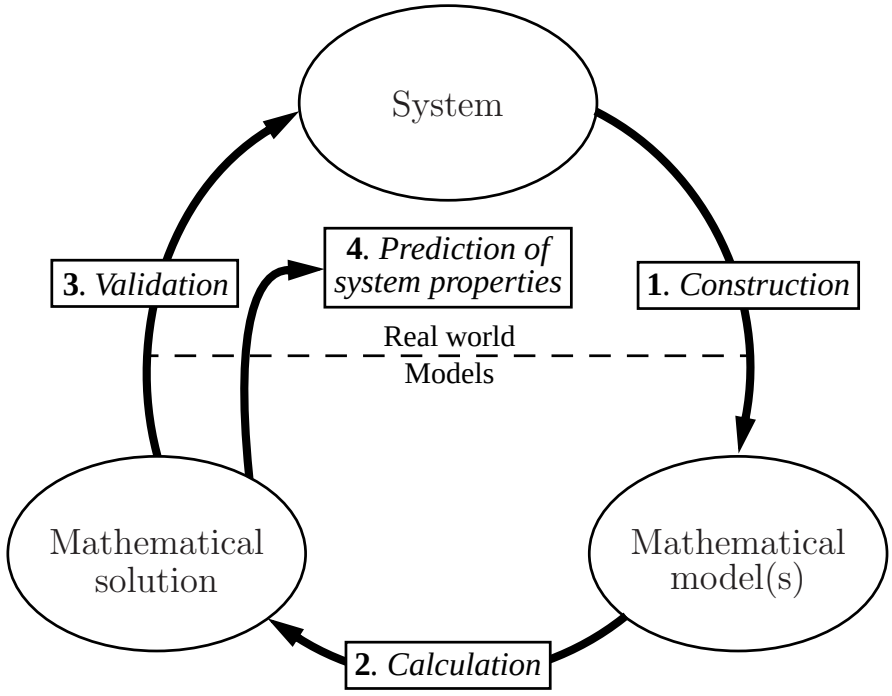

Figure 1.1: The mathematical modeling procedure.

1. Construction of the mathematical model or models

   Here, real quantities and processes of the system are replaced by a set of mathematical equations that describe the behavior of the system. The construction of a mathematical model typically includes various idealizations and approximations and, according to [69], may not be well-defined and frequently involves a high degree of creativity. The mathematical formulation should be consistent with the model use, meaning that any idealizations or approximations used are a consequence of the reasons for which the model is being developed. Because information is limited, the mathematical model for the system may not be unique, and more than one model may be considered.

2. Calculation of the mathematical solution

   Using analytical and/or numerical techniques, a set of equations is solved to find an approximation to the system properties. Unlike the previous step, rigorous procedures exist to do this, though some level of approximation may be required. This step may involve the creation of a computer model, *e.g.*, a finite element model. In the case of an approximate solution, a bound on the approximation error can usually be attained. We note that the mathematical solution may or may not be unique. Non-uniqueness may arise, for example, due to a collection of partial differential equations with no unique solution, or from inadequate solution techniques.

3. Validation of the model results

As defined by [75], p. 11, validation is the process of determining the degree to which the mathematical model is an accurate representation of the system. Data is required, and validation must include experimental observations of the system that were not used in model construction. There are rigorous validation procedures for a few applications under ideal conditions; attempts to generalize the validation process are underway [89].

4. Prediction of the system properties

   The validated mathematical model is used to make predictions about the properties of the system; these predictions are consistent with model use. For example, if the model use is design, we may use the predictions to assign values to certain design variables, based on the properties of the mathematical solution. Note that a prediction is made only after the validation step is complete.

Steps (2)-(4) have been, and continue to be, areas of active research; see [4, 75] for an extensive list of references on step (3). Model construction, step (1), has yet to be addressed in a rigorous manner and, as a result, is largely a subjective discipline; Kapur [61], p. 13, states that constructing models "is an art." Herein, we attempt to replace some of the subjective nature of model construction through the use of model selection, a rigorous mathematical procedure to select the optimal model from a collection of candidate models for the system. The elements of model construction are summarized in Section 1.2; model selection will be discussed in detail in Section 1.3.

1.2 Model construction

There are three philosophies for model construction (see Fig. 1.2):

1. Observation-based

   In this method, only data, the first type of available information, is used for model construction. Observations of the system are used to fit the properties of the model to those of the system, *i.e.*, curve-fitting procedures are used. This is typical of so-called nontechnical systems (biological, economic, sociological, *etc.*) where cause and effect relationships are unknown.

2. Physics-based

   In this method, only prior knowledge, the second type of available information, is used for model construction. The properties of the system are decomposed into subsystems whose behavior is assumed to follow known mathematical relations [68], p. 16. The dependencies among subsystems are then modeled and, in this manner, the cause and effect relationships of the system are quantified.

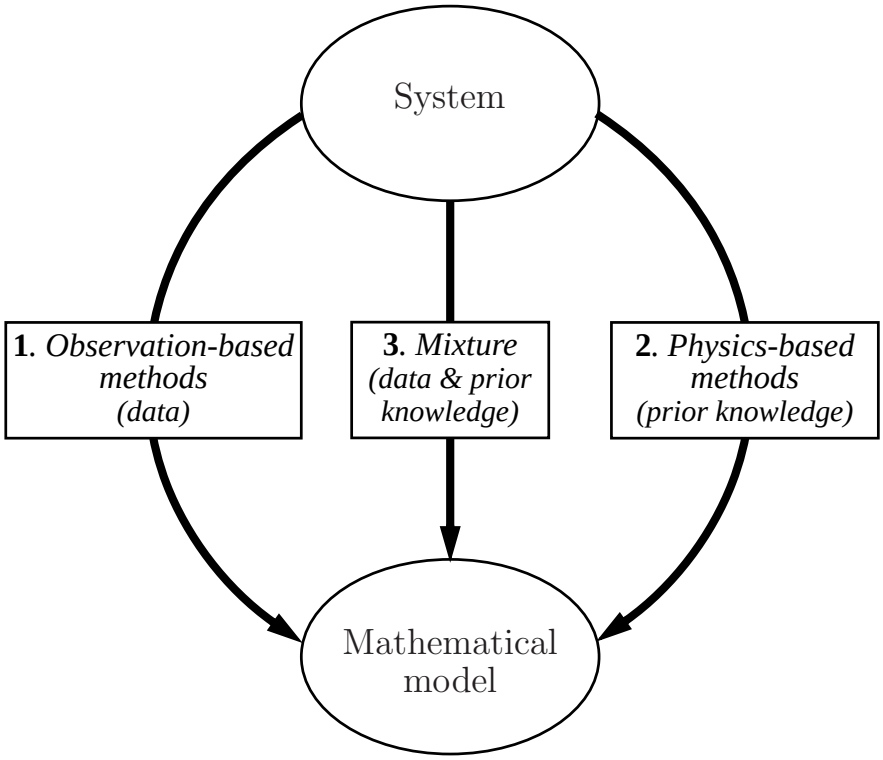

Figure 1.2: Philosophies for model construction.

3. Mixture

   Here, both data and prior knowledge on the system are utilized, and a mixture of the observation-based and physics-based approaches are used for model construction. This philosophy is typical of systems in engineering and applied science.

Example 1.1: We illustrate first the observation-based philosophy for model construction. Assume $n \geq 1$ observations of random variable Y are available. Based on the observations, we postulate a functional form (or a collection of possible functional forms) for the probability law for Y. Each functional form may contain a set of unknown parameters. We apply methods of parameter estimation (*e.g.*, the method of maximum likelihood [63]) to define the values of the parameters for each possible model form.

We next illustrate the physics-based philosophy for model construction. Assume that prior knowledge suggests Y is given by $Y = \exp(X)$, a continuous and hence measurable function of Gaussian random variable X. It can be shown that, under these assumptions, Y follows a lognormal distribution. $\Diamond$

As discussed in [67], there are three steps in the model construction process:

1. Identify important variables

All important variables to be included in the mathematical model are identified, *i.e.*, those variables supported by the available information, and pertinent to the intended model use. Because the available information is limited, some important variables can be overlooked, and other supercilious variables deemed important. The identification of important variables is often a subjective process; it is unrealistic to expect a general method for identifying important variables for all types of systems.

2. Define the functional form

   Our objective is to define a functional form for the equations that expresses how the variables interact to produce the effect or effects of interest. Examples include forms that are linear or nonlinear, discrete or continuous, and static or dynamic [37], p. 2. The functional form will most likely include some parameters. According to [18], p. 7, this process may also be subjective, which is why scientists and engineers must be trained, educated, and experienced in their discipline. In general, more than one set of equations will be consistent with the available information, meaning that the functional form for the model may not be unique.

3. Specify values for each model parameter

   Once a functional form for the model has been defined, the value for each model parameter must be specified. Parameters may be deterministic and/or stochastic. Because the available information is limited, more than one set of parameter values may be consistent with what is known, meaning that the values for the parameters may also be non-unique.

Regardless of the philosophy used to construct a mathematical model, the steps in the model construction process outlined above remain the same. Because the available information on a system is incomplete, **in general there can be no unique mathematical model for the system**.

Example 1.2: Let X be a random variable. Assume that the only available information on X are the second-moment properties: X has a mean of 3 and a standard deviation of 1/9. This is insufficient to uniquely define the probability law for X. Figure 1.3 illustrates four models for X, each consistent with the available information. In each case, histograms were generated with 10,000 independent samples. $\Diamond$

While hundreds of books and countless journal papers deal with estimation of model parameters (step (3)), relatively little has appeared concerning the specification of the mathematical form of the model [18], p. 5. This motivates model selection, a procedure to select the optimal model for the system from a class of candidate mathematical models.

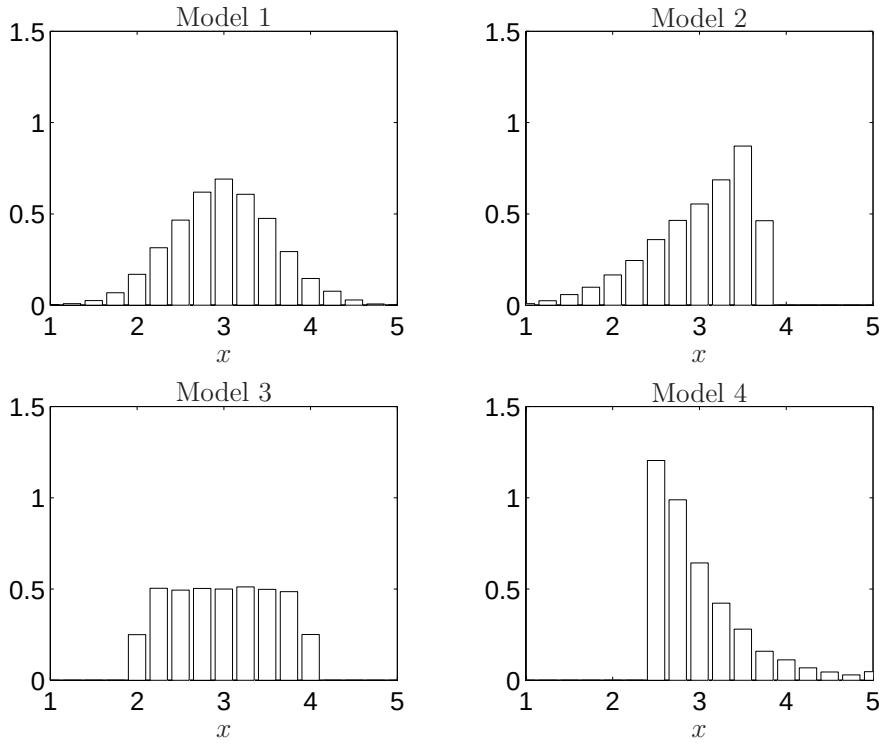

Figure 1.3: Four models for X, each with mean 3, standard deviation 1/9.

1.3 Model selection

The model selection process is illustrated by Fig. 1.4. As discussed, limited information about the system implies that the mathematical model for the system may not be unique. We therefore consider a class of candidate models for the system, where each model in the class is consistent with the available information. The objective is to select one model from the class of candidate models; we call this the optimal model. The optimal model depends on the available information, and the class of candidate models. The dashed line in Fig. 1.4 indicates that certain methods for model selection include the model use, while others do not. In the case of the former, the optimal model also depends on model use.

A model is a simplification of, or approximation for, reality (nature); a model is not truth by definition. It follows that, in general, none of the candidate models are an exact representation for nature. It is therefore a mistake to believe that there is a perfect model for the system that will somehow be uncovered in the model selection process [18], p. 11; we can only hope to select a model that provides an accurate approximation for nature. The model selection process attempts to rank models in the set of candidate models relative to each other; whether the optimal model is actually "good" depends on the quality of the science and *a priori* thinking that went into constructing each candidate model.

Methods for model selection can be applied to problems in one of two categories.

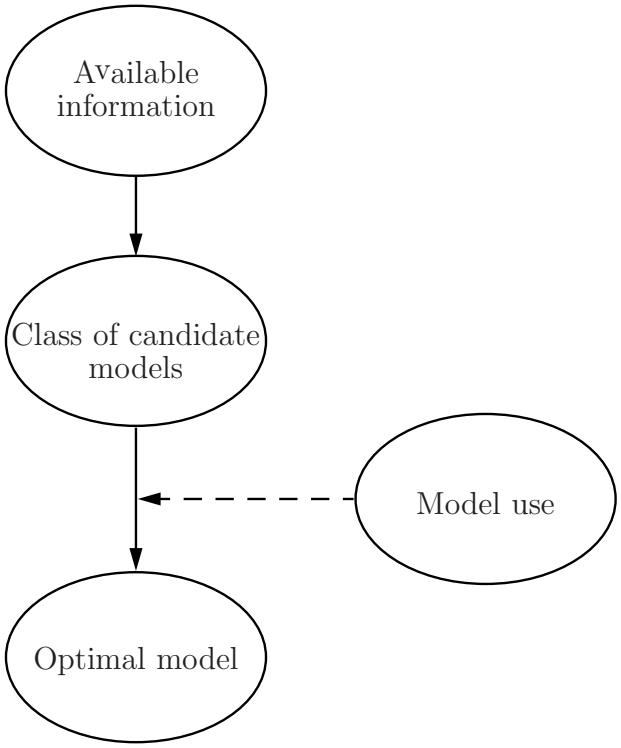

Figure 1.4: The model selection process.

- **Category I**: given a mathematical model for the system with a known functional form and a collection of unknown model parameters, we use methods for model selection to determine optimal values for the model parameters. Problems in this category are similar to problems in classical parameter estimation.
- **Category II**: given a collection of mathematical models for the system, where each is (possibly) of a different functional form with a collection of unknown parameters, we use methods for model selection to determine the optimal functional form for the model, as well as values for the model parameters.

Model selection is not a new concept; numerous methods exist (see [18], pp. 27-29) including, but not limited to:

1. Method of maximum likelihood ([35] and [63], Chapter 6)

   This technique was originally developed to provide point estimates for any unknown parameters of a given model form, *i.e.*, Category I problems; these estimates maximize a particular objective function, called the likelihood function. The method of maximum likelihood has since been extended for use with Category II problems to provide estimates of the likelihood of each model in a collection of candidate models. The candidate model with the largest likelihood is selected.

2. Bayesian methods [11, 98]

   As method (1), this method was originally developed for Category I problems. Given data from an experiment, a model with fixed functional form, and prior knowledge of a set of unknown parameters, this technique provides updated or posterior knowledge of the model parameters. It also has been extended to consider the more general Category II problems, where a collection of candidate models, each with a different functional form and a set of unknown parameters, is considered. In this case, the posterior knowledge gained from the Bayesian method includes a collection of posterior model probabilities, as well as posterior knowledge of the model parameters for each functional form; the model with the largest posterior model probability is then selected to study the system properties. A technique called Bayesian model averaging [80] may also be used, where the system properties are estimated not by one model, but by the average of a subclass of all models considered, weighted by the posterior model probabilities. We refer to methods (1) and (2) as classical methods for model selection.

3. Decision theory [81, 84]

   The decision-theoretic method includes a representation for possible models of the system, usually referred to as the possible states of nature, a set of options for action, and a utility function that quantifies the decision maker's preferences for each action taken, under each possible nature state. This method for model selection can be used to estimate model parameters, as well as to select the optimal functional form for a mathematical model from a collection of candidate models, *i.e.*, Category I and II problems, respectively. The decision-theoretic approach is more general than the classical methods since, with a special choice of utility function, the classical solution can be recovered.

   The form of the utility function is consistent with the model use. For example, consider two mathematical models for a system, A and B, where the models are to be used to design the system (model use). Suppose the consequences of using model A is preferential to the consequences of using model B for the purpose of design. We quantify this preference via an appropriate utility function; note that the form of the utility function may change under a different model use.

There are several advantages to the decision-theoretic method for model selection (see Table 1.1). First, classical methods for model selection, *i.e.*, methods (1) and (2), cannot be used unless data is available. In addition, these methods can be unreliable when data exists, but is limited [34]. The decision-theoretic approach does not have these limitations. Second, classical methods for model selection ignore model use, meaning that for a given set of data, the selected model for the system does not change under a different model use. In contrast, as we have seen, model use is included in the decision-theoretic approach to model selection through an appropriate utility function. This is especially important when constructing mathematical models for systems where the consequences of using an incorrect model can be disastrous,

Table 1.1. Methods for model selection.

Method	Data	Model use	Philosophy
Classical	Required	Ignored	Observation-based
Decision theory	Not required	Included	Any

e.g., the collapse of a bridge during an earthquake or the structural failure of a spacecraft during atmospheric re-entry. Herein, we refer to such systems as high risk. Third, classical methods for model selection are typically compatible only with the observation-based philosophy for model construction; this makes sense since it is under this philosophy that the methods were developed. The decision-theoretic approach, which makes use of both data and model use, is compatible with all three of the philosophies for model construction discussed in Section 1.2.

One criticism of the decision-theoretic approach to model selection is the possible subjectivity to the definition of the utility function. For example, consider a committee comprised of several experts; it is often the case that they differ on their definition of the utility function [84], p. 83. Hence, it is crucial to the decision-theoretic approach for model selection that we understand the behavior of the system, as well as the consequences of system failure, well enough to construct an appropriate utility function. For the applications considered here, we assume this is the case.

Model selection has been studied extensively in numerous applications in the areas of biology [18], meteorology [57], Section 7.3, finance [22, 70, 74], and geology [26]. Model construction for these types of systems typically follows the observation-based philosophy and is applied to Category I problems. Category II problems have been considered in a limited way, *e.g.*, linear regression with unknown order (functional form) and unknown coefficients (model parameters). With the exception of finance, applications are typically limited to simple models, *e.g.*, random variables or vectors. Methods (1)-(3) for model selection are routinely used for calculations.

There have been relatively few applications of model selection in applied science and engineering. Model construction for these types of systems typically follows a physics-based or mixed philosophy, and complex stochastic behavior is often present. Most of the applications in these fields use classical methods for calculation, require that data be abundantly available, and study is limited to simple models, *e.g.*, random variables or vectors. Applications include Bayesian model updating for structural dynamics and structural reliability [7, 8, 62, 76, 90, 97], Bayesian analysis of model form uncertainty [3], Appendix D, [58], and structural fatigue [101]. Very few applications use the decision-theoretic approach; applications include structural reliability [38, 46, 51], and engineering materials [92].

Herein, we use the decision-theoretic approach for model selection for systems:

(1) with little or no available data, and (2) that involve complex stochastic behavior, *i.e.*, non-Gaussian random vectors and stochastic processes. Special attention is given to the high risk system, where the consequences of using an inappropriate model can be disastrous. The applications considered are diverse; examples include the prediction of the response of an internal component in a spacecraft during atmospheric re-entry, and the design of a distributed sensor network for vehicle surveillance.

The model selection problem is defined in Chapter 2; methods for analysis are also discussed. Simple applications are used to illustrate many aspects of the model selection problem in Chapter 3. One class of stochastic model, the polynomial chaos approximation, is defined and studied under the context of model selection in Chapter 4. Two high risk applications, namely the atmospheric re-entry of a spacecraft and the design and analysis of a distributed sensor network for vehicle surveillance, are discussed in Chapters 5 and 6, respectively. Conclusions are discussed in Chapter 7. Several appendices are included.

This page intentionally left blank.

Chapter 2

The model selection problem

Generalities and essentials of the model selection problem are presented. A definition of the problem is presented in Section 2.1. The method of maximum likelihood and Bayesian methods, herein referred to as classical methods for model selection, are discussed in Section 2.2, while the decision-theoretic approach to model selection is discussed in Section 2.3. A series of related examples is used for illustration. A sensitivity analysis of these methods for model selection in discussed in Section 2.4.

2.1 Problem definition

A framework for mathematical modeling is presented in Section 2.1.1, and the steps of the model selection process, first introduced in Section 1.3, are discussed in detail in Section 2.1.2.

2.1.1 Framework for mathematical modeling

Most systems in science and engineering can be described by an input/output relationship of the type shown in Fig. 2.1, where input $\mathbf{Z}$ and operator $\mathcal{D}$ are, in general, vector-valued. Typically, $\mathcal{D}$ is defined by a collection of differential, integral, and/or algebraic equations with (possibly) random coefficients. The objective is to estimate properties of an output vector, $\mathbf{Y}$. For example, $\mathcal{D}$ can be a finite element model of a spacecraft that maps an applied pressure field, $\mathbf{Z}$, to the displacement response, $\mathbf{Y}$, of an internal component. Properties of $\mathbf{Y}$, *e.g.*, the maximum in time, can then be estimated. Note that our definition of a system includes $\mathbf{Z}$, $\mathcal{D}$, and $\mathbf{Y}$, as illustrated by the dashed lines in Fig. 2.1.

Models for $\mathbf{Z}$, and/or $\mathcal{D}$ may be random or deterministic and, in general, are defined by: (1) a functional form, and (2) a collection of parameters. Often, the

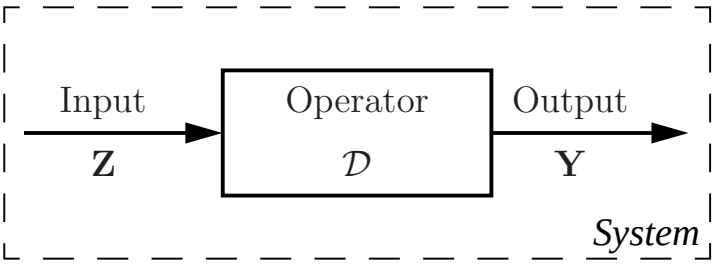

Figure 2.1: Model for a system as an input/output relationship.

information on the system is limited, meaning that the functional form for $\mathbf{Z}$ and/or $\mathcal{D}$, as well as their parameters, may not be known precisely. In this case, we can define a collection of candidate models for $\mathbf{Z}$, and/or $\mathcal{D}$, and use methods of model selection to choose the optimal members from the collection.

2.1.2 Steps of model selection process

The model selection process requires four steps:

1. gather the **available information** on the system,
2. define a **class of candidate models** for the system,
3. consider the **model use**, and
4. calculate the **optimal model**.

Each of these steps is discussed in detail in the following sections.

2.1.2.1 Available information

Denote by $\mathcal{I}$ the available information on $\mathbf{Z}$ and/or $\mathcal{D}$. As discussed in Section 1.1, $\mathcal{I}$ is of two types: (1) data and/or (2) prior knowledge. If nothing is known about the system behavior, we say the amount of available information is zero; if we have complete understanding of the system behavior, we say the amount of available information is complete. For applications, the amount of available information is nonzero, but less than complete; we call this **limited** or **incomplete information**.

2.1.2.2 Class of candidate models

Define

$$\mathcal{M} = \{\text{all models for } \mathbf{Z} \text{ and/or } \mathcal{D} \text{ consistent with } \mathcal{I}\}\,. \tag{2.1}$$

$\mathcal{M}$ has the following properties: (1) it is nonempty and may contain an uncountable number of models, and (2) the number of models in $\mathcal{M}$ is nonincreasing with increasing $\mathcal{I}$. A member, m, of $\mathcal{M}$ is a **candidate model** for $(\mathbf{Z}, \mathcal{D})$; each $m \in \mathcal{M}$ is consistent with $\mathcal{I}$.

2.1.2.3 Model use

Model use refers to the purpose for which estimates of the properties of $\mathbf{Y}$ are considered. Several examples were discussed in Section 1.1. Herein, we study two different types of model use in detail: (1) for **design**, *i.e.*, select $m \in \mathcal{M}$ so that the properties of $\mathbf{Y}$ satisfy some constraints, and (2) for **performance prediction**, *i.e.*, select $m \in \mathcal{M}$ to predict system performance. Note that in general, the estimates of the properties of $\mathbf{Y}$ are dependent on $\mathcal{I}$, $\mathcal{M}$, and model use.

2.1.2.4 Optimal model

Denote by $m^\star \in \mathcal{M}$, the **optimal model** for $\mathbf{Z}$ and $\mathcal{D}$. Three methods to select $m^\star$ will be discussed. Classical methods for model selection, which includes the method of maximum likelihood and Bayesian methods, will be discussed in Section 2.2; as will be shown, these methods are independent of model use. A decision-theoretic approach to model selection is presented in Section 2.3. Note that, in general, $m^\star \in \mathcal{M}$ depends on the:

1. available information, $\mathcal{I}$,
2. class of candidate models, $\mathcal{M}$, and
3. model use (decision-theoretic approach only).

The sensitivity of $m^\star$ to changes in (1), (2), and/or (3) is discussed in Section 2.4.

Because $\mathcal{I}$ is, in general, incomplete, there may be additional models for the system, not in $\mathcal{M}$, that are superior to all candidate models considered. This is consistent with Section 1.3, where we remark that it is a mistake to believe that there is a perfect model for the system that will somehow be uncovered in the model selection process. In fact, the use of $m^\star$ to model the system does not imply belief that the optimal model represents reality exactly; it implies only that under particular

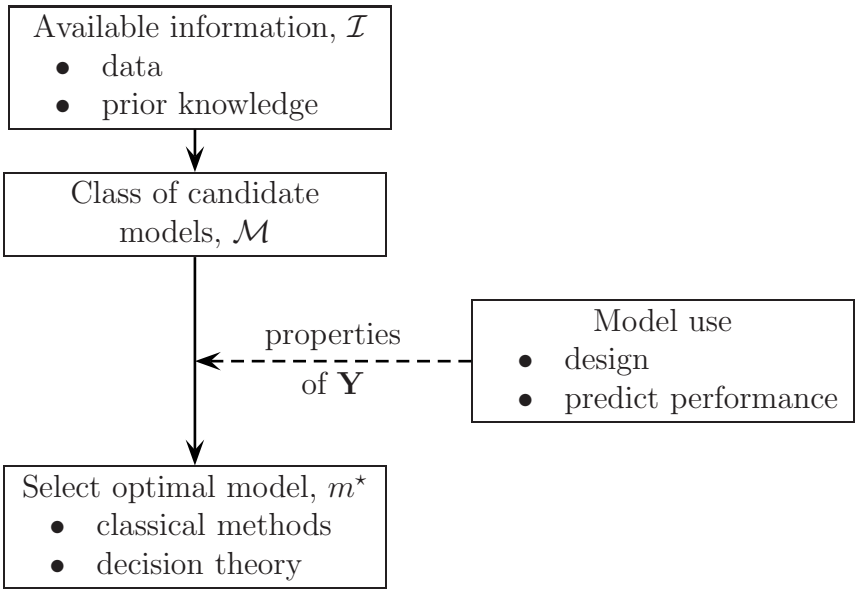

Figure 2.2: The model selection process.

circumstances (available information, resources, *etc.*) $m^\star$ is pragmatically optimal [51].

The model selection process is illustrated by Fig. 2.2; the dashed line is used to indicate that model use may or may not be considered. We next present an example to illustrate these definitions.

Example 2.1: Consider the following input/output relationship:

$$Y = \mathcal{D}(Z; x) = Z + x, \tag{2.2}$$

where x is a deterministic parameter, and Z is a random variable. Assume the information on the probability law for Z is incomplete, and we are interested in the following property of Y

$$P\left(Y \leq d\right), \tag{2.3}$$

where d is a specified critical value for Y. We consider two cases of model use: (1) design, and (2) performance prediction. For (1), x is a design parameter and we set its value to achieve some design constraint, *i.e.*, $P\left(Y \leq d\right) = \bar{q}$, where $\bar{q}$ is a specified level of reliability. For (2), the value for x is known and fixed.

The available information on Z is limited.

- Data: $n \geq 1$ measurements of Z, $\{z_1, z_2, \ldots, z_n\}$.
- Prior knowledge: the probability law (CDF) for Z is one of $F_i(z; \boldsymbol{\theta}_i)$, where $\boldsymbol{\theta}_i$ is a vector of unknown parameters, $i = 1, 2, \ldots$

Accordingly, let

$$\mathcal{M} = \{F_i(z; \boldsymbol{\theta}_i)\},\ i = 1, 2, \ldots, \tag{2.4}$$

be the class of all candidate models for Z. As discussed in Section 1.3, this is a Category II model selection problem. The objective is to select an optimal model, $m^\star \in \mathcal{M}$, for random variable Z; note that $m^\star$ under use (1) may be different than $m^\star$ under use (2). $\Diamond$

2.2 Classical methods for model selection

In this section, we discuss the classical methods for model selection. This includes the method of maximum likelihood, defined in Section 2.2.1 and Bayesian methods, defined in Section 2.2.2. As discussed in Chapter 1, classical methods for model selection typically ignore model use, require data, and can be unreliable when data is limited.

Define the available information, $\mathcal{I}$, to be given by independent samples of $\mathbf{Z}$ and $\mathcal{D}$, denoted $\{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_n\}$, and $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_q\}$, respectively, as well as any prior knowledge on $\mathbf{Z}$ and $\mathcal{D}$. We make two assumptions: (1) $\mathbf{Z}$ and $\mathcal{D}$ are independent, and (2) the prior knowledge on $\mathbf{Z}$ and $\mathcal{D}$ is sufficient to define a collection of absolutely continuous probability laws for $\mathbf{Z}$ and $\mathcal{D}$.

Let $F_{\mathbf{Z}}^{(i)}(\mathbf{z}; \boldsymbol{\theta}_i)$, $i = 1, 2, \ldots$, be a collection of candidate probability laws (CDFs) for $\mathbf{Z}$ with unknown parameter vector $\boldsymbol{\theta}_i$. Conditional on the functional form for $\mathcal{D}$, the remaining uncertainty in the operator relates to its coefficients. Accordingly, let $F_{\mathcal{D}}^{(j)}(\mathbf{x}; \boldsymbol{\phi}_j)$, $j = 1, 2, \ldots$, be the collection of candidate CDFs for the coefficients of $\mathcal{D}$, for all possible functional forms for $\mathcal{D}$; $\boldsymbol{\phi}_j$ is the corresponding unknown parameter vector. For example, consider the case of two candidate functional forms for the operator, $\mathcal{D}_1(X)$ and $\mathcal{D}_2(X)$, and three candidate CDFs for the random coefficient, X. In this case, we consider the collection $F_{\mathcal{D}}^{(j)}$, $j = 1, 2, \ldots, 6$. Let $f_{\mathbf{Z}}^{(i)}(\mathbf{z}; \boldsymbol{\theta}_i)$, $i = 1, 2, \ldots$, and $f_{\mathcal{D}}^{(j)}(\mathbf{x}; \boldsymbol{\phi}_j)$, $j = 1, 2, \ldots$, denote the corresponding candidate PDFs for $\mathbf{Z}$ and $\mathcal{D}$, respectively. Then

$$\mathcal{M} = \{m_{ij}\} = \left\{ f_{\mathbf{Z},\mathcal{D}}^{(ij)}(\mathbf{z}, \mathbf{x}; \boldsymbol{\theta}_i, \boldsymbol{\phi}_j) = f_{\mathbf{Z}}^{(i)}(\mathbf{z}; \boldsymbol{\theta}_i) \times f_{\mathcal{D}}^{(j)}(\mathbf{x}; \boldsymbol{\phi}_j) \right\},\ i, j = 1, 2, \ldots \tag{2.5}$$

is a collection of candidate models for $\mathbf{Z}$ and $\mathcal{D}$, where $f_{\mathbf{Z},\mathcal{D}}^{(ij)}$ is the joint PDF of $\mathbf{Z}$ and $\mathcal{D}$.

It remains to define the procedure to select the optimal model, $m^\star \in \mathcal{M}$. Classical methods for model selection are one way to do this. Using the method of maximum likelihood, discussed in Section 2.2.1, $m^\star$ is a single joint PDF for $\mathbf{Z}$ and $\mathcal{D}$, with point estimates for parameters $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$. With Bayes' method, discussed in Section 2.2.2, $m^\star$ is a single joint PDF for $\mathbf{Z}$ and $\mathcal{D}$, but the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ are treated as random vectors. Examples are presented to illustrate concepts.

2.2.1 The method of maximum likelihood

Recall the **likelihood function** for $\mathbf{Z}$ and $\mathcal{D}$ [63, 98]

$$l_{\mathbf{Z}}^{(i)}(\boldsymbol{\theta}_i) = \prod_{k=1}^{n} f_{\mathbf{Z}}^{(i)}(\mathbf{z}_k; \boldsymbol{\theta}_i),\ i = 1, 2, \ldots$$
$$l_{\mathcal{D}}^{(j)}(\boldsymbol{\phi}_j) = \prod_{r=1}^{q} f_{\mathcal{D}}^{(j)}(\mathbf{x}_r; \boldsymbol{\phi}_j),\ j = 1, 2, \ldots, \tag{2.6}$$

where $\mathbf{z}_k,\ k = 1, 2, \ldots, n$, and $\mathbf{x}_r,\ r = 1, 2, \ldots, q$, denote the available data. By assumption (1), we define

$$l_{ij}(\boldsymbol{\theta}_i, \boldsymbol{\phi}_j) = l_{\mathbf{Z}}^{(i)}(\boldsymbol{\theta}_i)\ l_{\mathcal{D}}^{(j)}(\boldsymbol{\phi}_j),\ i, j = 1, 2, \ldots \tag{2.7}$$

to be the joint likelihood function for $\mathbf{Z}$ and $\mathcal{D}$. We interpret l_{ij} to be the likelihood that model m_{ij} is true. Note that $l_{ij} < \infty$, for $i, j \geq 1$ since it is the product of a finite number of PDFs. We can therefore define a collection of normalized likelihood functions, given by

$$\bar{l}_{ij} = \frac{l_{ij}}{\sum_{i,j=1}^{\infty} l_{ij}} \tag{2.8}$$

where each $\bar{l}_{ij}$ takes values in $[0, 1]$.

To select $m^\star \in \mathcal{M}$, we follow a two-step process:

1. select optimal values for parameter vectors $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$, and
2. select optimal functional forms for PDFs $f_{\mathbf{Z}}$ and $f_{\mathcal{D}}$.

For step (1), the maximum likelihood point estimates, $\hat{\boldsymbol{\theta}}_i$ and $\hat{\boldsymbol{\phi}}_j$, of parameter vectors $\boldsymbol{\theta}_i$ and $\boldsymbol{\phi}_j$, $i, j = 1, 2, \ldots$, are the values for $\boldsymbol{\theta}_i$ and $\boldsymbol{\phi}_j$ that maximize Eq. (2.7). Assuming vectors $\boldsymbol{\theta}_i$ and $\boldsymbol{\phi}_j$ have lengths r_i and s_j, respectively, $i, j = 1, 2, \ldots$, $\hat{\boldsymbol{\theta}}_i$ and $\hat{\boldsymbol{\phi}}_j$ solve

$$\frac{\partial l_{ij}(\boldsymbol{\theta}_i, \boldsymbol{\phi}_j)}{\partial \theta_{i,k}} = 0,\ k = 1, 2, \ldots, r_i,$$
$$\frac{\partial l_{ij}(\boldsymbol{\theta}_i, \boldsymbol{\phi}_j)}{\partial \phi_{j,k}} = 0,\ k = 1, 2, \ldots, s_j, \tag{2.9}$$

such that

$$\frac{\partial^2 l_{ij}(\boldsymbol{\theta}_i, \boldsymbol{\phi}_j)}{\partial \theta_{i,k}^2} < 0,\ k = 1, 2, \ldots, r_i,$$
$$\frac{\partial^2 l_{ij}(\boldsymbol{\theta}_i, \boldsymbol{\phi}_j)}{\partial \phi_{j,k}^2} < 0,\ k = 1, 2, \ldots, s_j. \tag{2.10}$$

For step (2), we select $m^\star = \left(f_{\mathbf{Z}}^{(i)}(\mathbf{z}; \hat{\boldsymbol{\theta}}_i), f_{\mathcal{D}}^{(j)}(\mathbf{x}; \hat{\boldsymbol{\phi}}_j) \right) \in \mathcal{M}$ such that

$$\bar{l}_{ij}(\hat{\boldsymbol{\theta}}_i, \hat{\boldsymbol{\phi}}_j) \geq \bar{l}_{kr}(\hat{\boldsymbol{\theta}}_k, \hat{\boldsymbol{\phi}}_r), \ k, r = 1, 2, \ldots \tag{2.11}$$

To illustrate the method of maximum likelihood for model selection, we consider two examples. In the first, the probability law for random variable Z is known and we seek optimal values for parameter vector $\boldsymbol{\theta}$; this is a Category I problem. In the second example, a Category II problem, the probability law for random variable Z is known to be one of a collection of candidate probability laws, each with unknown parameters; we seek the optimal probability law and its associated parameter vector.

Example 2.2: Assume the following information on random variable Z is available.

- Data: n independent samples of random variable Z, $\{z_1, z_2, \ldots, z_n\}$.
- Prior knowledge: Z is a Gaussian random variable.

By Eq. (2.6), the likelihood function is

$$l(\mu, \sigma) = \prod_{k=1}^{n} f(z_k; \mu, \sigma) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\sum_{k=1}^{n} \frac{(z_k - \mu)^2}{2\sigma^2} \right], \tag{2.12}$$

where

$$f(z; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2} \left(\frac{z - \mu}{\sigma} \right)^2 \right] \tag{2.13}$$

is the PDF of a Gaussian random variable, and $\boldsymbol{\theta} = (\mu, \sigma)^T$ are unknown parameters. By Eq. (2.9), we have

$$\begin{aligned} \frac{\partial(\ln l)}{\partial \mu} &= -\frac{1}{\sigma^2} \sum_{k=1}^{n} (z_k - \mu) = 0, \text{ and} \\ \frac{\partial(\ln l)}{\partial \sigma} &= -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{k=1}^{n} (z_k - \mu)^2 = 0. \end{aligned} \tag{2.14}$$

The solution to Eq. (2.14) gives $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\sigma})^T$, the maximum likelihood estimate of $\boldsymbol{\theta}$, *i.e.*, [63], p. 124

$$\begin{aligned} \hat{\mu} &= \frac{1}{n} \sum_{k=1}^{n} z_k, \text{ and} \\ \hat{\sigma}^2 &= \frac{1}{n} \sum_{k=1}^{n} (z_k - \hat{\mu})^2. \end{aligned} \tag{2.15}$$

The solution can be interpreted as the optimal model over the space of parameters $(\mu, \sigma > 0)$. $\Diamond$

Example 2.3: Next assume the available information on random variable Z is given by the following.

- Data: n independent samples of random variable Z, $\{z_1, z_2, \ldots, z_n\}$.
- Prior knowledge: Z is either a Gaussian or exponential random variable.

The candidate models for Z are $\mathcal{M} = \{f^{(i)}(z; \hat{\boldsymbol{\theta}}_i)\}$, $i = 1, 2$, where $f^{(1)}$ is f from Eq. (2.13), and

$$f^{(2)}(z; \alpha, \nu) = \nu \exp\left[-\nu(z - \alpha)\right], \ z \geq \alpha. \tag{2.16}$$

is the PDF for an exponential random variable. $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_2$ are maximum likelihood point estimates of the unknown parameter vectors $\boldsymbol{\theta}_1 = (\mu, \sigma)^T$, and $\boldsymbol{\theta}_2 = (\alpha, \nu)^T$.

By step (1), we need to calculate $\hat{\boldsymbol{\theta}}_i$, $i = 1, 2$; Eq. (2.15) gives $\hat{\boldsymbol{\theta}}_1$. Using Eq. (2.9), we have

$$\begin{aligned} \hat{\alpha} &= \min_j z_j, \text{ and} \\ \hat{\nu} &= \frac{n}{\sum_{k=1}^{n}(z_k - \hat{\alpha})}. \end{aligned} \tag{2.17}$$

Equation (2.6) then gives values for $l_i(\hat{\boldsymbol{\theta}}_i)$, $i = 1, 2$.

Step (2) is to select $m^\star = f^{(i)}(z; \hat{\boldsymbol{\theta}}_i)$, $i = 1$ or 2. Consider the case where $n = 5$ samples are available, *i.e.*,

$$\{z_1, z_2, \ldots, z_5\} = \{4.14, 3.37, 4.50, 3.01, 5.66\}.$$

The maximum likelihood point estimates for the parameters are $\hat{\mu} = 4.14$, $\hat{\sigma}^2 = 0.860$, $\hat{\alpha} = 3.01$, and $\hat{\nu} = 0.887$; by Eq. (2.8), $\bar{l}_1 = 0.234$ and $\bar{l}_2 = 0.766$. Hence, by Eq. (2.11), $m^\star = f^{(2)}(z; \hat{\alpha}, \hat{\nu})$. $\Diamond$

2.2.2 Bayesian methods

Consider a collection of mutually exclusive events $H_1, H_2, \ldots, H_r \in \mathcal{F}$, defined on probability space $(\Omega, \mathcal{F}, P)$, where $\cup_{i=1}^{r} H_i = \Omega$ and $H_i \cap H_j = \emptyset$ for $i \neq j$. In this case, we say $\{H_i\}$, $i = 1, 2, \ldots, r$, forms a partition of Ω. For event $E \in \mathcal{F}$, Bayes' theorem states [31], p. 124

$$P(H_j \mid E) = \frac{P(E \mid H_j)P(H_j)}{\sum_{i=1}^{r} P(E \mid H_i)P(H_i)}, \ j = 1, 2, \ldots, r. \tag{2.18}$$

Often, we write Eq. (2.18) as a relation of proportionality, *i.e.*,

$$P(H_j \mid E) \propto P(E \mid H_j)P(H_j), \ j = 1, 2, \ldots, r. \tag{2.19}$$

If we interpret H_1, H_2, ..., H_r as possible hypotheses, and E as evidence gained from performing an experiment, then Eq. (2.18) dictates how opinions on these hypotheses held before the experiment should be modified after the experiment takes place [85], p. 80.

In terms of the available information on the system, $\mathcal{I}$, the experimental evidence is the collection of independent samples $\{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_n\}$, and $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_q\}$. The hypotheses, H_j, $j = 1, 2, \ldots, r$, are any prior knowledge on $\mathbf{Z}$ and $\mathcal{D}$; this prior knowledge is quantified by:

1. **prior probabilities**, p'_{ij}, $i, j = 1, 2, \ldots$, that $m_{ij} \in \mathcal{M}$ is the correct model for the system, and
2. **prior PDFs** $f'^{(ij)}_{\boldsymbol{\Theta},\boldsymbol{\Phi}}(\boldsymbol{\theta}_i, \boldsymbol{\phi}_j)$, $i, j = 1, 2, \ldots$, of the parameter vectors, $\boldsymbol{\Theta}$ and $\boldsymbol{\Phi}$.

We require the prior knowledge to satisfy the following conditions:

$$\sum_{i,j=1}^{\infty} p'_{ij} = 1, \text{ and} \tag{2.20a}$$

$$\int_{\mathbb{D}_{\boldsymbol{\theta}_i}} \int_{\mathbb{D}_{\boldsymbol{\phi}_j}} f'^{(ij)}_{\boldsymbol{\Theta},\boldsymbol{\Phi}}(\boldsymbol{\theta}_i, \boldsymbol{\phi}_j) \, \mathrm{d}\boldsymbol{\phi}_j \, \mathrm{d}\boldsymbol{\theta}_i = 1, \; i, j = 1, 2, \ldots, \tag{2.20b}$$

where $\mathbb{D}_{\boldsymbol{\theta}_i}$ and $\mathbb{D}_{\boldsymbol{\phi}_j}$ denote the domain of vectors $\boldsymbol{\theta}_i$ and $\boldsymbol{\phi}_j$, respectively.

Direct application of Bayes' theorem, Eq. (2.19), gives

$$f''^{(ij)}_{\boldsymbol{\Theta},\boldsymbol{\Phi}}(\boldsymbol{\theta}_i, \boldsymbol{\phi}_j) \propto f'^{(ij)}_{\boldsymbol{\Theta},\boldsymbol{\Phi}}(\boldsymbol{\theta}_i, \boldsymbol{\phi}_j) l_{ij}(\boldsymbol{\theta}_i, \boldsymbol{\phi}_j), \; i, j = 1, 2, \ldots, \tag{2.21}$$

where $f''^{(ij)}_{\boldsymbol{\Theta},\boldsymbol{\Phi}}(\boldsymbol{\theta}_i, \boldsymbol{\phi}_j)$, $i, j = 1, 2, \ldots$, is a collection of **posterior PDFs** for $\boldsymbol{\Theta}$ and $\boldsymbol{\Phi}$, and $l_{ij}(\boldsymbol{\theta}_i, \boldsymbol{\phi}_j)$ is the joint likelihood function defined by Eq. (2.7). The **posterior probability**, p''_{ij}, that $m_{ij} \in \mathcal{M}$ is the correct model for the system can be computed [51]

$$p''_{ij} \propto p'_{ij} \int_{\mathbb{D}_{\boldsymbol{\theta}_i}} \int_{\mathbb{D}_{\boldsymbol{\phi}_j}} f'^{(ij)}_{\boldsymbol{\Theta},\boldsymbol{\Phi}}(\boldsymbol{\theta}_i, \boldsymbol{\phi}_j) l_{ij}(\boldsymbol{\theta}_i, \boldsymbol{\phi}_j) \, \mathrm{d}\boldsymbol{\phi}_j \, \mathrm{d}\boldsymbol{\theta}_i, \; i, j = 1, 2, \ldots \tag{2.22}$$

The posterior knowledge must also satisfy some conditions. First, the posterior probabilities are normalized, *i.e.*, Eq. (2.20a) with p'_{ij} replaced by p''_{ij}. Second, each posterior PDF is scaled to integrate to one, *i.e.*, Eq. (2.20b) with $f'^{(ij)}_{\boldsymbol{\Theta},\boldsymbol{\Phi}}$ replaced by $f''^{(ij)}_{\boldsymbol{\Theta},\boldsymbol{\Phi}}$.

Under the Bayesian method, the optimal model is then $m^\star = m_{ij} \in \mathcal{M}$ such that

$$p''_{ij} \geq p''_{kr}, \; k, r = 1, 2, \ldots \tag{2.23}$$

Note that with the Bayesian method, we do not approximate the unknown parameter vectors $\mathbf{\Theta}$ and $\mathbf{\Phi}$ using point estimates; we instead model them as absolutely continuous random vectors.

To illustrate the Bayesian method for model selection, we consider the same two examples from Section 2.2.1, but here we use the Bayesian method to select optimal models for random variable Z. In the first example, the probability law for random variable Z is known and we seek an optimal model for parameter vector $\boldsymbol{\theta}$; this is a Category I problem. In the second example, a Category II problem, the probability law for random variable Z is known to be one of a collection of candidate probability laws, each with unknown parameters; we seek the optimal probability law and a model for its associated parameter vector.

Example 2.4: This example is taken from [98], pp. 14-17. Assume the following information is available on random variable Z.

- Data: n independent samples of random variable Z, $\{z_1, z_2, \ldots, z_n\}$.
- Prior knowledge: (1) Z is Gaussian, (2) $\text{Var}[Z] = \sigma^2$, and (3) Z has unknown mean μ, with prior PDF $f'(\mu)$, where

$$f'(\mu) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left[-\frac{1}{2}\left(\frac{\mu - \mu_1}{\sigma_1}\right)^2\right], \tag{2.24}$$

  and parameters μ_1 and σ_1^2 are known and fixed.

By Eq. (2.21), the posterior PDF for the mean of Z is given by

$$\begin{aligned} f''(\mu) &\propto f'(\mu) l(\mu) \\ &\propto \exp\left[-\frac{1}{2}\left(\frac{1}{\sigma^2}\sum_{k=1}^{n}(z_k - \mu)^2\right) + \left(\frac{\mu - \mu_1}{\sigma_1}\right)^2\right], \end{aligned} \tag{2.25}$$

where the likelihood function, $l(\mu)$ given by Eq. (2.12), is written as a function of μ alone since here the value for σ is known. We then scale $f''(\mu)$ such that $\int_{-\infty}^{\infty} f''(\mu)\, \mathrm{d}\mu = 1$. The solution can be interpreted as the optimal model over the space of parameter μ. $\Diamond$

Example 2.5: Next assume the available information on Z is given by the following.

- Data: n independent samples taken from random variable Z.
- Prior knowledge: (1) $\text{Var}[Z] = 2$, (2) Z is either a Gaussian or exponential random variable with equal probability, *i.e.*, $p_1' = p_2' = 1/2$, (3) if Z is Gaussian,

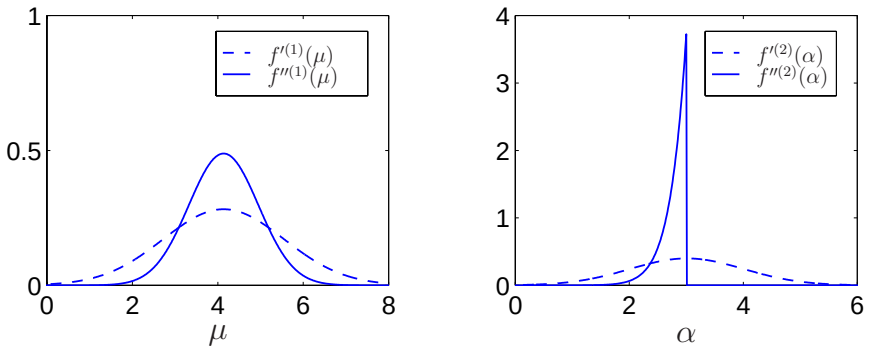

Figure 2.3: Prior and posterior PDFs for μ (left) and α (right).

parameter σ^2 is known; the unknown parameter μ, has prior PDF, $f'^{(1)}(\mu)$, given by Eq. (2.24), and (4) if Z is exponential, parameter ν is known; the unknown parameter α, has prior PDF, $f'^{(2)}(\alpha)$, given by

$$f'^{(2)}(\alpha) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left[-\frac{1}{2}\left(\frac{\alpha - \mu_2}{\sigma_2}\right)^2\right], \tag{2.26}$$

where parameters μ_2 and σ_2^2 are known and fixed.

The candidate models for Z are $\mathcal{M} = \{f^{(i)}(z;\Theta_i)\}$, $i = 1, 2$, where $f^{(1)}$ is the PDF of a Gaussian random variable, given by Eq. (2.13), with $\Theta_1 = \mu$, and $f^{(2)}$ is the PDF of an exponential random variable, given by Eq. (2.16), with $\Theta_2 = \alpha$.

The posterior PDF for μ is given by Eq. (2.25); the posterior PDF for α can be computed via

$$\begin{aligned} f''^{(2)}(\alpha) &\propto f'^{(2)}(\alpha) l_2(\alpha) \\ &\propto \exp\left[-\nu \sum_{k=1}^{n} (z_k - \alpha) - \frac{1}{2}\left(\frac{\alpha - \mu_2}{\sigma_2}\right)^2\right]. \end{aligned} \tag{2.27}$$

We then scale $f''^{(i)}$, $i = 1, 2$, such that $\int_{-\infty}^{\infty} f''^{(i)}(\theta_i)\, \mathrm{d}\theta_i = 1$. Equation (2.22) gives the posterior probability that Z is Gaussian or exponential. With parameters $\sigma^2 = 2$, $\nu = \sqrt{2}/2$, $\mu_1 = \hat{\mu}$, $\sigma_1^2 = 2$, $\mu_2 = \hat{\alpha}$, $\sigma_2^2 = 1$, and the samples from Example 2.3, we have $p_1'' = 0.993$ and $p_2'' = 0.007$; thus the Gaussian model is optimal. The prior and posterior PDFs for μ and α are shown in Fig. 2.3. $\lozenge$

2.3 Decision-theoretic method for model selection

Modeling in applied science and engineering involves the making of decisions [51]. It is therefore natural to consider the model selection process outlined in Section 2.1 as

a problem in decision theory, herein referred to as the **decision-theoretic method for model selection**. In contrast to the classical methods for model selection studied in Section 2.2, the decision-theoretic method for model selection includes model use, does not require data and is reliable when data is limited, and allows us to quantify the consequences of choosing one model of the system over another. We present an overview of decision theory in Section 2.3.1, then formulate the model selection problem as a problem in decision theory in Section 2.3.2.

2.3.1 Decision theory

Consider an event for which we have incomplete information, called the **state of nature**; because the information on the state of nature is limited, it is uncertain. Suppose further we must choose an action to take, but the consequences of taking an action depend on the state of nature. For example, consider the problem of deciding whether or not to bring an umbrella to school, given that the state of nature for the day (the weather) is uncertain. Decision theory is a rigorous mathematical framework for making the optimal decision from a collection of candidate choices, under the uncertainty in the state of nature. It is a special case of Game theory [13, 93], where the game is played against nature [21], p. 16. We begin the discussion with a general framework for problems in decision theory, and then present a simple example to illustrate concepts.

The general problem in decision theory has four ingredients [21], p. 10:

1. the collection of **candidate actions**, $\mathcal{A}$,
2. the collection of all **possible states of nature**, $\mathcal{M}$,
3. the **utility function**, U, and
4. the **optimal action**, $a^\star \in \mathcal{A}$.

Each of these ingredients is discussed in detail in the following sections.

2.3.1.1 Candidate actions

A decision problem exists when there is a choice of candidate actions to take; the consequence of these actions must depend on the state of nature [21], p. 10. Define

$$\mathcal{A} = \{\text{all candidate actions}\} = \{a_i\},\ i = 1, 2, \ldots \tag{2.28}$$

to be the collection of candidate actions. By construction, $\mathcal{A}$ is countable.

2.3.1.2 Possible states of nature

Recall that $\mathcal{I}$ denotes any information on the state of nature. Define

$$\mathcal{M} = \{\text{all possible states of nature consistent with } \mathcal{I}\} = \{m_j\},\ j \in J, \tag{2.29}$$

where J is an index set that, in general, is uncountable. Note that $m_j \in \mathcal{M}$, $j \in J$, is consistent with the available information, $\mathcal{I}$.

Let p_j denote the probability that m_j is true. We require $\sum_{j \in J} p_j = 1$. Note that $p_j = 1$ implies the state of nature is known to be m_j, *i.e.*, $\mathcal{I}$ is complete, and $p_j = 0$ implies m_j is not a possible state of nature. Hence, without loss of generality we consider the case where $p_j \in (0, 1)$, $\forall j \in J$.

2.3.1.3 Utility

Let the mapping

$$(a_i, m_j) \mapsto c_{ij},\ i = 1, 2, \ldots,\ \forall j \in J, \tag{2.30}$$

define the **consequence** for action, a_i, assuming the state of nature is m_j. Let U denote the utility function; U takes nonnegative values and quantifies the consequences of each action, under each state of nature, *i.e.*, $U \colon \mathcal{A} \times \mathcal{M} \to [0, \infty)$.

For some problems, it is possible to define U directly; in the general case, we use the consequences, c_{ij}, to assign properties for U. Assume the consequences can be ordered according to their preference ([27], Chapter 12, and [93]). To illustrate, consider, for consequences c_{ij}, c_{kl}, c_{pq}, and c_{rs}, the following preference ordering:

$$c_{ij} \prec c_{kl} \preceq c_{pq} \sim c_{rs}. \tag{2.31}$$

We say consequence c_{kl} is **preferred** to consequence c_{ij}, consequence c_{pq} is **preferred or equivalent** to consequence c_{kl}, and consequence c_{rs} is **strictly equivalent** to consequence c_{pq}. We assume the preference ordering is transitive, *i.e.*,

$$c_{ij} \prec c_{kl} \prec c_{pq} \Longrightarrow c_{ij} \prec c_{pq}. \tag{2.32}$$

The utility is defined such that it satisfies the following conditions:

$$\begin{aligned} U(a_i, m_j) < U(a_{i'}, m_{j'}) &\iff c_{ij} \succ c_{i'j'}, \\ U(a_i, m_j) \leq U(a_{i'}, m_{j'}) &\iff c_{ij} \succeq c_{i'j'},\ \text{and} \\ U(a_i, m_j) = U(a_{i'}, m_{j'}) &\iff c_{ij} \sim c_{i'j'}. \end{aligned} \tag{2.33}$$

Note that, with this construction, U has the property of scale indifference [54], p. 159, *i.e.*,

$$\begin{aligned} U(a_i, m_j) \leq U(a_{i'}, m_{j'}),\ i = 1, 2, \ldots,\ \forall j \in J \Longrightarrow \\ \hat{U}(a_i, m_j) \leq \hat{U}(a_{i'}, m_{j'}),\ i = 1, 2, \ldots,\ \forall j \in J, \end{aligned} \tag{2.34}$$

where $\hat{U} = \alpha U + \beta$, and $\alpha > 0$, β are deterministic constants. Further, the utility function may not be unique.

In the literature, it is usual to: (1) construct U such that the complement of the LHS of Eq. (2.33) is true, *i.e.*, $U(a_i, m_j) > U(a_{i'}, m_{j'}) \iff c_{ij} \succ c_{i'j'}$, *etc.*, (2) define the opportunity loss as $L = -U$, and (3) use L for calculations [84], p. 60. With our construction, we have avoided this step.

2.3.1.4 Optimal action

By definition, the utility, U, is a random variable; we define the **expected utility** of action a_i

$$u(a_i) = \mathrm{E}[U(a_i, \mathcal{M})] = \sum_{j \in J} U(a_i, m_j)\ p_j,\ \ i = 1, 2, \ldots, \tag{2.35}$$

where p_j, defined in Section 2.3.1.2, is the probability that state of nature m_j is true. Action a_k is optimal if and only if $u(a_i)$ is a minimum for $i = k$ [46]; in this case, we denote a_k by $a^\star$, where

$$u(a^\star) \leq u(a_i),\ \ i = 1, 2, \ldots \tag{2.36}$$

It follows that $a^\star$ is the most preferable action, *i.e.*, $a^\star \succeq a_i$, $i = 1, 2, \ldots$.

Example 2.6: Given that it may rain today in Ithaca, should Tyler: (1) bring an umbrella to school today, or (2) leave it at home? The possible states of nature are

$$\begin{aligned} m_1 &= \{\text{it rains today}\},\ \text{and} \\ m_2 &= \{\text{it doesn't rain today}\}. \end{aligned} \tag{2.37}$$

Tyler can estimate p_j, $j = 1, 2$, in one of several ways. First, by using data, *e.g.*, he can build a time series model from weather observations in Ithaca over the last 50 years and use it to forecast the weather for today. Second, he can use any prior knowledge, *e.g.*, from watching the Weather Channel. Third, he can use "expert" opinion, *e.g.*, he can ask his dad.

Tyler's candidate actions are

$$\begin{aligned} a_1 &= \{\text{bring the umbrella}\},\ \text{and} \\ a_2 &= \{\text{don't bring the umbrella}\}, \end{aligned} \tag{2.38}$$

and his preferences define the ordering of the consequences, c_{ij}, of action a_i, under state of nature m_j. Assume he assigns the following ordering

$$c_{21} \prec c_{12} \prec c_{11} \prec c_{22}, \tag{2.39}$$

		Nature	
		m_1	m_2
Action	a_1	1	2
	a_2	5	0

Figure 2.4: Utility function, $U(a_i, m_j)$, for example decision problem.

meaning that the consequence of leaving the umbrella at home and it rains, c_{21}, is least preferable because he has to sit in class all day with wet clothes; by Eq. (2.39), all other consequences are preferable to this one.

Using Eq. (2.33), Tyler can define values for the utility of action a_i, when the state of nature is m_j; this will not uniquely define U. Suppose he decides on the values listed in Fig. 2.4; the expected utility of each action can be calculated by Eq. (2.35)

$$u(a_i) = \mathrm{E}[U(a_i, \mathcal{M})] = U(a_i, m_1)\ p_1 + U(a_i, m_2)\ p_2,\ i = 1, 2. \tag{2.40}$$

With $p_1 = 2/3$ and $p_2 = 1/3$, Eq. (2.40) gives $u(a_1) = 4/3$, $u(a_2) = 10/3$. The optimal action for Tyler is therefore $a^\star = a_1$ by Eq. (2.36). ◊

2.3.1.5 Model selection uncertainty

Consider the scenario when the form of the utility function is uncertain; this can occur, for example, if a committee comprised of several experts with differing opinions is asked to define the utility. In addition, recall Fig. 2.4 used in Example 2.6. The values for U were assigned from an adult point of view; most likely a child is not as concerned with having wet clothes and would select a different utility.

We assume: (1) we can express the utility function in a general functional form, *e.g.*, a Fourier expansion, with a vector of coefficients given by $\boldsymbol{\beta}$; in this case we write $U(\mathcal{A}, \mathcal{M}; \boldsymbol{\beta})$, and (2) the uncertainty in the utility can be completely described by allowing the coefficients, $\boldsymbol{\beta}$, to be a random vector with known probability law, F. Under these assumptions, Eq. (2.35) is a conditional expected utility; it follows that [84], p. 83

$$u(a_i) = \int_{\mathbb{D}_{\boldsymbol{\beta}}} \sum_{j \in J} U(a_i, m_j; \boldsymbol{\beta})\ p_j\ F(\mathrm{d}\boldsymbol{\beta}),\ i = 1, 2, \ldots \tag{2.41}$$

is the unconditional expected utility of action $a_i \in \mathcal{A}$, where $\mathbb{D}_{\boldsymbol{\beta}}$ denotes the domain of vector $\boldsymbol{\beta}$.

2.3.2 Application of decision theory to model selection

Recall Fig. 2.1, where the objective is to select models for input, $\mathbf{Z}$, and/or operator, $\mathcal{D}$, to estimate properties of output, $\mathbf{Y}$, for a particular model use. As we have seen,

in general the available information on the system we wish to model is incomplete. As a result, models for $\mathbf{Z}$ and/or $\mathcal{D}$ will exhibit some degree of inconsistency with the true behavior of the system, and estimates of the properties of $\mathbf{Y}$ will be incorrect. We seek a procedure by which to choose models for $\mathbf{Z}$ and/or $\mathcal{D}$ such that we achieve conservative estimates of the properties of $\mathbf{Y}$; our definition of "conservative" will depend on the model use. The decision-theoretic method for model selection is one way to achieve this. We present a general formulation of the method in Section 2.3.2.1, and then apply the method for two types of model use: (1) for design, discussed in Section 2.3.2.2, and (2) for performance prediction, discussed in Section 2.3.2.3.

2.3.2.1 General formulation

Let $\mathcal{A} = \{a_1, a_2, \ldots, a_q\}$ denote the collection of candidate actions; practical limitations require $\mathcal{A}$ to be finite. The character of the objects contained in $\mathcal{A}$ reflect the model use. For example, if the model use is design, each $a_i \in \mathcal{A}$ is a candidate design for the system (or some collection of subsystems); if the model use is performance prediction, each $a_i \in \mathcal{A}$ is a candidate model for the system (or some collection of subsystems).

$\mathcal{M}$ is the complete, *i.e.*, exhaustive, uncountable collection of models for the system, where each $m \in \mathcal{M}$ is consistent with the available information on the system. For calculations, we consider a finite subcollection

$$\mathcal{M}' = \{m_1, m_2, \ldots, m_n\} \subseteq \mathcal{M}, \tag{2.42}$$

with corresponding probabilities p_i, $i = 1, 2, \ldots, n$, that m_i is true. The collections $\mathcal{A}$ and $\mathcal{M}'$ do not coincide, in general; $\mathcal{A} = \mathcal{M}'$ is a special case.

It is desirable, for some $\epsilon > 0$ small, to have

$$\sum_{j=1}^{n} p_j > 1 - \epsilon, \tag{2.43}$$

meaning that the probability of any model being true in $\mathcal{M} \setminus \mathcal{M}'$ is small. In the general case, there is no way to verify the models in $\mathcal{M}'$ satisfy Eq. (2.43). Further, we normalize p_i such that $\sum_{i=1}^{n} p_i = 1$. For each $i = 1, 2, \ldots, n$, p_i can be estimated by: (1) the likelihood of model m_i (Section 2.2.1), (2) the Bayesian posterior probability of m_i (Section 2.2.2), or (3) in the absence of data, non-informative prior probabilities or "expert" opinion.

A utility function is used to quantify the consequences of taking action a_i, if model m_j is true. Define $U \colon \mathcal{A} \times \mathcal{M}' \to [0, \infty)$ as

$$U(a_i, m_j) = \gamma(a_i) + \psi(a_i, m_j), \ i = 1, 2, \ldots, q, \ j = 1, 2, \ldots, n, \tag{2.44}$$

		Model, $\mathcal{M}' \subseteq \mathcal{M}$	
		m_j	
Action, $\mathcal{A}$	a_i	$U(a_i, m_j)$	

Figure 2.5: Utility function, $U(a_i, m_j)$, for model selection problem.

where $\gamma(a_i) \geq 0$ is the **cost** of action a_i, and $\psi(a_i, m_j) \geq 0$ is the **penalty** of action a_i, if model m_j is true. It follows that the expected utility of each $a_i \in \mathcal{A}$ is given by

$$u(a_i) = \sum_{j=1}^{n} U(a_i, m_j)\ p_j = \gamma(a_i) + \sum_{j=1}^{n} \psi(a_i, m_j)\ p_j,\ i = 1, 2, \ldots, q. \tag{2.45}$$

The optimal action, $a^\star \in \mathcal{A}$, minimizes Eq. (2.45). The concept of utility is further illustrated for the general case by the array in Fig. 2.5. Note that if we choose not quantify the consequences of taking a particular action, it is equivalent to assessing U for a single cell in the array.

To illustrate the cost, γ, and penalty, ψ, recall the values for the utility function from Example 2.6, listed in Fig. 2.4. The cost of bringing the umbrella is one, and the cost of leaving the umbrella at home is zero. Hence, $\gamma(a_1) = 1$ and $\gamma(a_2) = 0$. The diagonal elements of the array in Fig. 2.4 are assigned zero penalty since, for these cases, Tyler guesses correctly, *i.e.*, he brings the umbrella if it rains, and leaves it at home if it does not. The off-diagonal elements of the array correspond to an incorrect guess; penalties $\psi(a_1, m_2) = 2$ and $\psi(a_2, m_1) = 4$ are assigned accordingly. In the general case, the functional form for γ and ψ depend on model use; we will define both for the cases of design and performance prediction in Sections 2.3.2.2 and 2.3.2.3, respectively.

Example 2.7: Consider the special case, discussed in [46], p. 23, where:

1. $\mathcal{A} = \mathcal{M}'$, and
2. $U(a_i, m_j) = 1 - \delta_{ij}$, where $\delta_{ij} = 1$ for $i = j$ and zero otherwise.

The expected utility of each candidate action is given by

$$u(a_i) = \sum_{j=1}^{n} (1 - \delta_{ij})\ p_j = 1 - p_i,\ i = 1, 2, \ldots, q, \tag{2.46}$$

so that $a^\star = m_k$ if $p_k \geq p_i$, $i = 1, 2, \ldots, n$. Hence, the optimal model under the decision-theoretic method for model selection is identical to the optimal model under the classical methods for model selection discussed in Section 2.2; if we assign p_j to be the likelihood of m_j, the maximum likelihood solution is recovered, while if we assign p_j to be the posterior probability of m_j, the Bayesian solution is recovered. $\Diamond$

2.3.2.2 Model use: design

Suppose the objective is to construct a mathematical model for the system so that the properties of the output, $\mathbf{Y}$, satisfy some constraints; we call this a **design problem**. We first develop a general formulation for the utility function introduced in Section 2.3.2.1 assuming the model use is design, then present an example for illustration.

Consider the case when $r \geq 1$ properties of $\mathbf{Y} \in \mathbb{R}^{d'}$ are of interest. Denote by $g_k(\mathbf{Y})$, $k = 1, 2, \ldots, r$, the properties of $\mathbf{Y}$, where each $g_k \colon \mathbb{R}^{d'} \to \mathbb{R}^r$ is a deterministic, measurable mapping. For example, $g_1(\mathbf{Y}) = \max_j Y_j$ and $g_2(\mathbf{Y}) = \sum_j Y_j$ define two properties of $\mathbf{Y}$, where Y_j denotes the j-th coordinate of $\mathbf{Y}$. The objective is then to select a design for the system such that, for $k = 1, 2, \ldots, r$, we have $P(g_k(\mathbf{Y}) \leq d_k) = \bar{q}_k$, where d_k and $\bar{q}_k$, are deterministic parameters that define the design constraints. However, the probability law for $\mathbf{Y}$ is not known because, in the general case, information on $\mathbf{Y}$ is incomplete.

In the case of model use for design, each member of $\mathcal{A}$ is a candidate design for the system. The objective of the design problem can be stated as follows: select design $a_i \in \mathcal{A}$ such that, for $j = 1, 2, \ldots, n$, and $k = 1, 2, \ldots, r$, we have

$$P\left(g_k(\mathbf{Y}) \leq d_k \mid a_i, m_j\right) = \bar{q}_k. \tag{2.47}$$

Note that the probability law for $\mathbf{Y}$ under design a_i and model m_j is perfectly known, so that $P\left(g_k(\mathbf{Y}) \leq d_k \mid a_i, m_j\right)$ can be calculated. Further, the condition in Eq. (2.47) may not be satisfied, *i.e.*,

$$P\left(g_k(\mathbf{Y}) \leq d_k \mid a_i, m_j\right) \neq \bar{q}_k, \text{ some } k \in \{1, 2, \ldots, r\}. \tag{2.48}$$

Recall the definition of the utility function, given by Eq. (2.44), where $\gamma(a_i)$ denotes the cost of design a_i, and $\psi(a_i, m_j)$ denotes the penalty associated with design a_i, if model m_j is true. Let

$$\gamma(a_i) = \text{ cost of implementing design } a_i. \tag{2.49}$$

For example, we set $\gamma(a_i) > \gamma(a_j) \geq 0$ if design a_i requires more material, funds, development time, *etc.*, than design a_j.

Consistent with Eq. (2.30), we need a method to order the consequences of design a_i, assuming m_j is true, for each $i = 1, 2, \ldots, q$ and $j = 1, 2, \ldots, n$, in order to define the penalty function, $\psi(a_i, m_j)$. Let vector $\boldsymbol{\xi}_{ij} \in \mathbb{R}^r$ be such that

$$P\left(g_k(\mathbf{Y}) \leq \xi_{ij,k} \mid a_i, m_j\right) = \bar{q}_k, \ k = 1, 2, \ldots, r, \tag{2.50}$$

where $\xi_{ij,k}$ denotes the k-th coordinate of $\boldsymbol{\xi}_{ij}$. It follows that

$$\xi_{ij,k} \leq d_k \iff P\left(g_k(\mathbf{Y}) \leq d_k \mid a_i, m_j\right) \geq \bar{q}_k, \text{ and} \tag{2.51a}$$

$$\xi_{ij,k} > d_k \iff P\left(g_k(\mathbf{Y}) \leq d_k \mid a_i, m_j\right) < \bar{q}_k. \tag{2.51b}$$

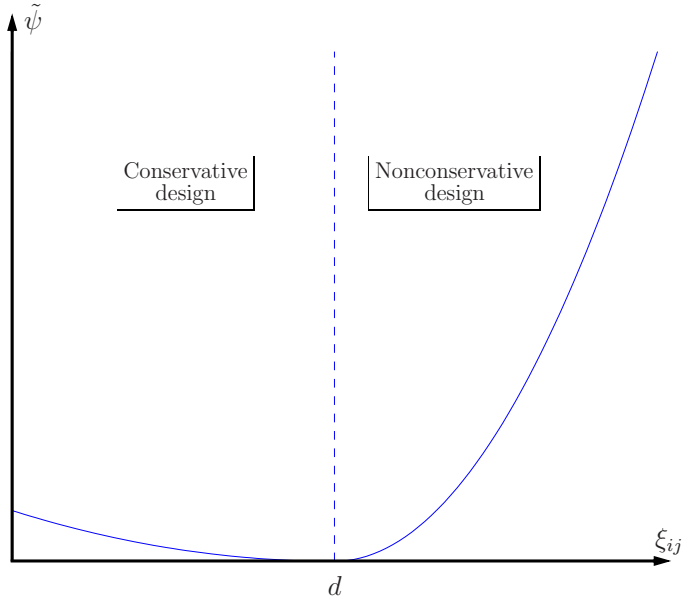

Figure 2.6: Example penalty function for design.

If a_i satisfies Eq. (2.51a), we say it is a **conservative design** with respect to property k, while if a_i satisfies Eq. (2.51b), we say it is a **non-conservative design** with respect to property k. A conservative design is preferable to a non-conservative design; we use Eqs. (2.51) to order the consequences of each (a_i, m_j) pair.

For example, consider the case where the objective is to constrain one property of scalar random variable, Y, *i.e.*, $r = d' = 1$; Eq. (2.51) reduces to

$$\begin{aligned} \xi_{ij} \leq d &\iff P\left(g(Y) \leq d \mid a_i, m_j\right) \geq \bar{q}, \text{ and} \\ \xi_{ij} > d &\iff P\left(g(Y) \leq d \mid a_i, m_j\right) < \bar{q}. \end{aligned} \tag{2.52}$$

An appropriate penalty function is given by

$$\psi(a_i, m_j) = \tilde{\psi}(\xi_{ij}, d) = \begin{cases} \beta_1(\xi_{ij} - d)^2 & \text{if } \xi_{ij} \leq d, \text{ and} \\ \beta_2(\xi_{ij} - d)^2 & \text{if } \xi_{ij} > d, \end{cases} \tag{2.53}$$

where $\beta_1, \beta_2 \geq 0$ are deterministic parameters, and we replace ψ with $\tilde{\psi}$ to denote that the penalty assigned to design a_i when model m_j is true can be expressed as a function of ξ_{ij} and d alone. We require $\beta_2 > \beta_1$ since $\{\xi_{ij} \leq d\}$ is preferable to $\{\xi_{ij} > d\}$. $\tilde{\psi}$ is illustrated by Fig. 2.6, where it is evident that non-conservative designs are heavily penalized; note that a design that is overly conservative is also subject to a penalty.

The generalization to the case of $d', r > 1$ is straightforward. Define

$$\psi(a_i, m_j) = \tilde{\psi}(\boldsymbol{\xi}_{ij}, \mathbf{d}) = s(\boldsymbol{\xi}_{ij}, \mathbf{d}) \, \|\boldsymbol{\xi}_{ij} - \mathbf{d}\|^2, \tag{2.54}$$

where $\|\mathbf{y}\|$ denotes the 2-norm of vector $\mathbf{y}$, $\mathbf{d} \in \mathbb{R}^r$ is a vector with coordinates d_k,

$k = 1, 2, \ldots, r$, and $s(\boldsymbol{\xi}_{ij}, \mathbf{d})$ is a scaling function, given by

$$s(\boldsymbol{\xi}_{ij}, \mathbf{d}) = \sum_{k=1}^{2^r} \beta_k \; 1\left(\boldsymbol{\xi}_{ij} \in C_k(\mathbf{d})\right). \tag{2.55}$$

Each $\beta_k \geq 0$, $k = 1, 2, \ldots, 2^r$, is a deterministic parameter, and the collection $C_k(\mathbf{d}) \subseteq \mathbb{R}^r$, $k = 1, 2, \ldots, 2^r$, is given by

$$\begin{aligned}
C_1(\mathbf{d}) &= \underset{l=1}{\overset{r}{\times}} (-\infty, d_l] \\
C_2(\mathbf{d}) &= \left[\underset{\substack{l=1 \\ l \neq r}}{\overset{r}{\times}} (-\infty, d_l] \right] \times \left[\underset{\substack{l=1 \\ l \neq 1,2,\ldots,r-1}}{\overset{r}{\times}} (d_l, \infty) \right] \\
C_3(\mathbf{d}) &= \left[\underset{\substack{l=1 \\ l \neq r-1}}{\overset{r}{\times}} (-\infty, d_l] \right] \times \left[\underset{\substack{l=1 \\ l \neq 1,2,\ldots,r-2,r}}{\overset{r}{\times}} (d_l, \infty) \right] \\
&\vdots \\
C_{2^r}(\mathbf{d}) &= \underset{l=1}{\overset{r}{\times}} (d_l, \infty)
\end{aligned} \tag{2.56}$$

For example, with $r = 1$, Eq. (2.54) gives Eq. (2.53). With $r = 2$, Eq. (2.54) gives

$$\begin{aligned}
\tilde{\psi}(\boldsymbol{\xi}_{ij}, \mathbf{d}) = [\beta_1 1\left(\boldsymbol{\xi}_{ij} \in C_1(\mathbf{d})\right) + \beta_2 1\left(\boldsymbol{\xi}_{ij} \in C_2(\mathbf{d})\right) + \\
\beta_3 1\left(\boldsymbol{\xi}_{ij} \in C_3(\mathbf{d})\right) + \beta_4 1\left(\boldsymbol{\xi}_{ij} \in C_4(\mathbf{d})\right)] \left\|\boldsymbol{\xi}_{ij} - \mathbf{d}\right\|^2,
\end{aligned} \tag{2.57}$$

where

$$\begin{aligned}
C_1(\mathbf{d}) &= (-\infty, d_1] \times (-\infty, d_2], \\
C_2(\mathbf{d}) &= (-\infty, d_1] \times (d_2, \infty), \\
C_3(\mathbf{d}) &= (d_1, \infty) \times (-\infty, d_2], \\
C_4(\mathbf{d}) &= (d_1, \infty) \times (d_2, \infty).
\end{aligned} \tag{2.58}$$

If $\boldsymbol{\xi}_{ij} \in C_1(\mathbf{d})$, then design a_i under model m_j is conservative with respect to both $g_1(\mathbf{Y})$ and $g_2(\mathbf{Y})$, while if $\boldsymbol{\xi}_{ij} \in C_4(\mathbf{d})$, a_i under model m_j is non-conservative with respect to both properties. Suppose $\boldsymbol{\xi}_{ij} \in C_2(\mathbf{d})$ is preferable to $\boldsymbol{\xi}_{ij} \in C_3(\mathbf{d})$; we therefore select values for the coefficients of Eq. (2.57) such that $\beta_1 < \beta_2 < \beta_3 < \beta_4$. The scaling function, $s(\boldsymbol{\xi}_{ij}, \mathbf{d})$, is shown in Fig. 2.7, for $\beta_1 = 1$, $\beta_2 = 2$, $\beta_3 = 3$, and $\beta_4 = 5$.

The functional form of the utility used in this section, Eq. (2.54), provides the desired effect, *i.e.*, penalizes non-conservative designs, but it is not unique; alternative forms can be used. Regardless, the utility defined by Eq. (2.54) is sufficient for all design problems considered herein.

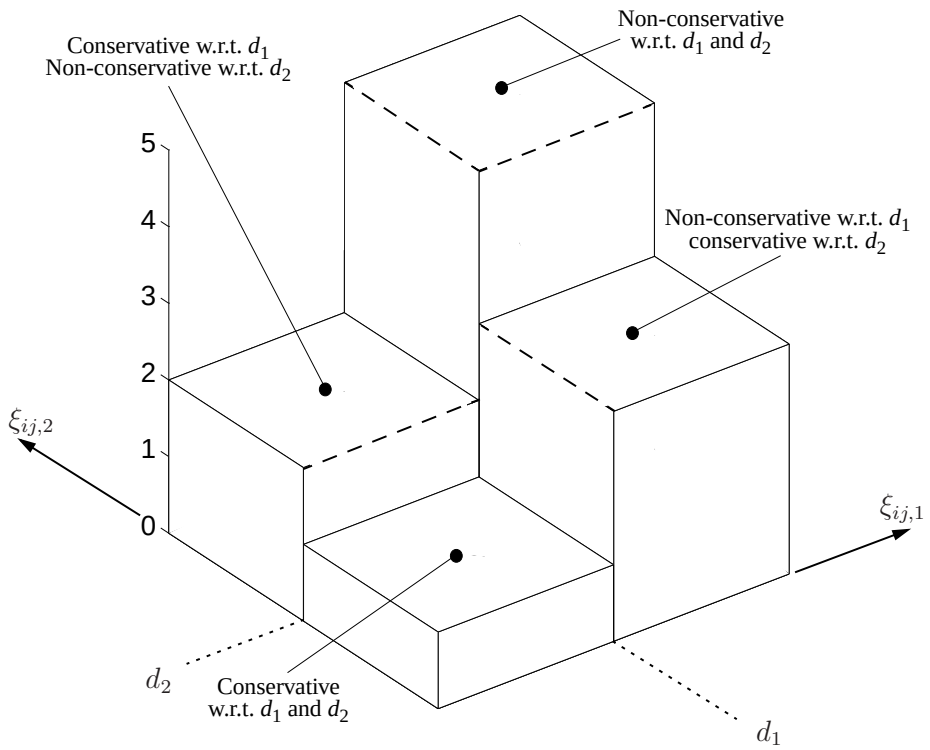

Figure 2.7: The function $s(\boldsymbol{\xi}_{ij}, \mathbf{d})$ for $r = 2$.

Example 2.8: Recall Example 2.1, where $Y = Z + x$. The objective is to select the value for x such that

$$P\left(Y \leq d\right) = \bar{q}, \tag{2.59}$$

under limited information on Z, where $\bar{q} \in [0, 1]$ is the required reliability of the design, and d is a specified critical value for Y. Note that, in this case, the property of interest is Y itself, *i.e.*, $g(Y) = Y$.

Assume the available information on Z is given by the following.

- Data: n independent samples of random variable Z.
- Prior knowledge: Z is either a Gaussian or lognormal random variable.

The candidate models for Z are given by $\mathcal{M}' = \{F_1(z; \hat{\boldsymbol{\theta}}_1), F_2(z; \hat{\boldsymbol{\theta}}_2)\}$, where F_1 is the CDF of a Gaussian random variable, F_2, given by

$$F_2(z; \boldsymbol{\theta}_2) = \int_0^z \frac{1}{\sqrt{2\pi}\zeta s} \exp\left[-\frac{1}{2}\left(\frac{\ln s - \eta}{\zeta}\right)^2\right] \mathrm{d}s, \tag{2.60}$$

is the CDF of a lognormal random variable, and $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_2$ are the maximum likelihood estimates of parameter vectors $\boldsymbol{\theta}_1 = (\mu, \sigma)^T$ and $\boldsymbol{\theta}_2 = (\eta, \zeta)^T$, respectively.

We consider two candidate designs, denoted by $\mathcal{A} = \{x_1, x_2\}$, where

$$x_i = d - F_i^{-1}(\bar{q}; \hat{\boldsymbol{\theta}}_i), \ i = 1, 2. \tag{2.61}$$

Let the cost of design x_i be quantified by

$$\gamma(x_i) = \gamma_0 x_i^2, \tag{2.62}$$

where $\gamma_0 \geq 0$ is a deterministic parameter. For design x_i under model m_j, the LHS of Eq. (2.59) is

$$P(Y \leq d \mid a_i, m_j) = P(Z \leq d - x_i \mid m_j) = F_j(d - x_i; \hat{\boldsymbol{\theta}}_j), \tag{2.63}$$

illustrating that, for $i \neq j$, the design constraint, Eq. (2.59), is violated. An appropriate penalty is Eq. (2.53), where we define ξ_{ij} such that

$$F_j(\xi_{ij} - x_i; \hat{\boldsymbol{\theta}}_j) = \bar{q}. \tag{2.64}$$

To illustrate, consider the identical $n = 5$ samples from Example 2.3, and let parameters $d = 6$, $\bar{q} = 0.99$, $\beta_1 = 1$, $\beta_2 = 10$, and $\gamma_0 = 1$. By Eq. (2.61), $x_1 = -0.296$ and $x_2 = -0.756$. Under (x_1, F_2), $\xi_{12} = x_1 + F_2^{-1}(\bar{q}; \hat{\boldsymbol{\theta}}_2) = 6.46$, so design x_1 is non-conservative if model F_2 is true by Eq. (2.51b). Likewise, under (x_2, F_1), $\xi_{21} = x_2 + F_1^{-1}(\bar{q}; \hat{\boldsymbol{\theta}}_1) = 5.54$, so design x_2 is conservative if model F_1 is true by Eq. (2.51a). Note that, by Eqs. (2.61) and (2.64), $\xi_{11} = \xi_{22} = d$. Using $p_1 = \bar{l}_1 = 0.453$ and $p_2 = \bar{l}_2 = 0.547$, by Eq. (2.45) we have

$$\begin{aligned} u(x_1) &= \gamma_0 x_1^2 + 0 \cdot p_1 + \beta_2(\xi_{12} - d)^2 p_2 = 1.25, \text{ and} \\ u(x_2) &= \gamma_0 x_2^2 + \beta_1(\xi_{21} - d)^2 p_1 + 0 \cdot p_2 = 0.667, \end{aligned} \tag{2.65}$$

so that design x_2 is optimal. $\Diamond$

2.3.2.3 Model use: predict performance

Next suppose the objective is to construct a mathematical model to assess the performance of the system; we call this a **prediction problem**. We first develop a general formulation for the utility function assuming the model use is performance prediction, then present an example for illustration.

Recall from Section 2.3.2.1 that $\mathcal{M}$ is the collection of all possible models for the system, and $\mathcal{M}' = \{m_1, m_2, \ldots, m_n\} \subseteq \mathcal{M}$ is a finite subcollection. In the case of model use for performance prediction, $\mathcal{A} = \{a_1, a_2, \ldots, a_q\} \subseteq \mathcal{M}$, so that each member of $\mathcal{A}$ is also a candidate model for the system. Most often, we have $\mathcal{A} = \mathcal{M}'$; in general, $\mathcal{A}$ and $\mathcal{M}'$ do not coincide. To illustrate the latter, consider the case where each member of $\mathcal{A}$ is an approximation for a member of $\mathcal{M}'$. This is one example where $\mathcal{A}$ and $\mathcal{M}'$ do not coincide, and is considered in Chapter 4.

As in Section 2.3.2.2, let $g_k(\mathbf{Y})$, $k = 1, 2, \ldots, r$, denote the properties of $\mathbf{Y}$ that are of interest, and let d_k, $k = 1, 2, \ldots, r$, be a collection of deterministic parameters. The objective is to assess the performance of the system, which we assume can be quantified by calculating, for all $k = 1, 2, \ldots, r$, values for $P(g_k(\mathbf{Y}) \leq d_k)$. Because information on $\mathbf{Y}$ is incomplete, the probability law for $\mathbf{Y}$ is not known.

Every $a_i \in \mathcal{A}$ and every $m_j \in \mathcal{M}'$ define a probability law for $\mathbf{Y}$. Let

$$P\left(g_k(\mathbf{Y}) \leq d_k \mid a_i\right),\ i = 1, 2, \ldots, q,\ k = 1, 2, \ldots, r, \tag{2.66}$$

denote a prediction of system performance using model a_i, and let

$$P\left(g_k(\mathbf{Y}) \leq d_k \mid m_j\right),\ j = 1, 2, \ldots, n,\ k = 1, 2, \ldots, r, \tag{2.67}$$

denote the system performance assuming m_j is true. Note that the probability law for $\mathbf{Y}$ under models a_i or m_j is perfectly known, so that the probabilities in Eqs. (2.66) and (2.67) can be calculated for $i = 1, 2, \ldots, q$, $j = 1, 2, \ldots, n$; this is true for the general case where $\mathcal{A}$ and $\mathcal{M}'$ do not coincide.

Assuming model m_j is true, if $a_i \neq m_j$, the prediction of system performance by Eq. (2.66) is inconsistent with Eq. (2.67). If model $a_i \in \mathcal{A}$ is such that

$$P\left(g_k(\mathbf{Y}) \leq d_k \mid a_i\right) \leq P\left(g_k(\mathbf{Y}) \leq d_k \mid m_j\right), \tag{2.68}$$

we say that model a_i results in a **conservative prediction** of performance with respect to property k; a_i results in a **non-conservative prediction** of performance with respect to property k if the complement of Eq. (2.68) is true. A conservative prediction is preferable to a non-conservative prediction; we use Eq. (2.68) to order the consequences of each (a_i, m_j) pair.

As defined by Eq. (2.44), the utility function has two parts: (1) the cost function, γ, and (2) the penalty function, ψ. In the case of model use for performance prediction, we have

$$\gamma(a_i) = \text{ cost of using model } a_i \text{ for prediction.} \tag{2.69}$$

For example, we set $\gamma(a_i) > \gamma(a_j) \geq 0$ if a prediction with model a_i requires additional effort, *e.g.*, CPU time, than a prediction with model a_j. We use $\psi(a_i, m_j)$ to penalize all $a_i \in \mathcal{A}$ that give non-conservative predictions of performance, *i.e.*, those models that violate Eq. (2.68). Let $\mathbf{v}_i \in \mathbb{R}^r$ be a vector with coordinates given by Eq. (2.66), and $\mathbf{h}_j \in \mathbb{R}^r$ be a vector with coordinates given by Eq. (2.67). Then

$$\psi(a_i, m_j) = \tilde{\psi}(\mathbf{v}_i, \mathbf{h}_j) = s(\mathbf{v}_i, \mathbf{h}_j)\ \|\mathbf{v}_i - \mathbf{h}_j\|^2, \tag{2.70}$$

where $s(\mathbf{v}_i, \mathbf{h}_j)$ is defined by Eq. (2.55) with $\boldsymbol{\xi}_{ij}$ and $\mathbf{d}$ replaced by $\mathbf{v}_i$ and $\mathbf{h}_j$, respectively. As in Eq. (2.54), we replace ψ with $\tilde{\psi}$ to denote that the penalty assigned to prediction under model a_i when model m_j is true can be expressed as a function of $\mathbf{v}_i$ and $\mathbf{h}_j$ alone. The penalty function defined by Eq. (2.70) is illustrated

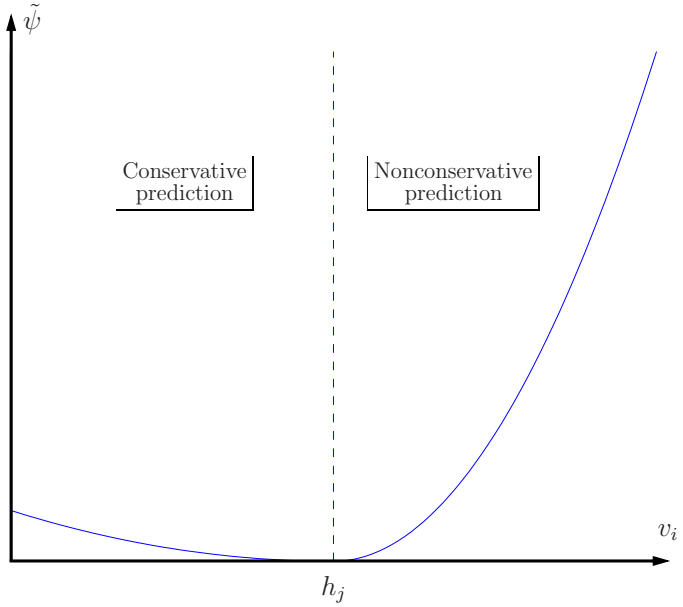

Figure 2.8: Example penalty function for prediction.

by Fig. 2.8 for the case of $r = 1$. Non-conservative predictions are heavily penalized; note that a prediction that is overly conservative is also subject to a penalty. For the special case when $\mathcal{A} = \mathcal{M}'$, we have $\mathbf{v}_i = \mathbf{h}_i$, $i = 1, 2, \ldots, n$, so that $\psi(m_i, m_j) = \tilde{\psi}(\mathbf{h}_i, \mathbf{h}_j)$.

The functional form of the utility used in this section, Eq. (2.70), provides the desired effect, *i.e.*, penalizes non-conservative predictions, but it is not unique; alternative forms can be used. Regardless, the utility defined by Eq. (2.70) is sufficient for all prediction problems considered herein.

Example 2.9: Recall Example 2.8, where we now assume the design to be fixed such that $Y = Z$ a.s., *i.e.*, $x = 0$. The objective is to predict $P(Y \leq d)$ under the same limited information on Z, where d is a specified critical value for Y. Let $\mathcal{A} = \mathcal{M}'$, where $\mathcal{M}'$ is the same as in Example 2.8, and denote by m_1 and m_2 the members $\mathcal{M}'$. The cost of a prediction with either model in $\mathcal{M}'$ is identical, so we set $\gamma(m_1) = \gamma(m_2) = 0$.

A prediction with model m_i is given by

$$h_i = P(Y \leq d \mid m_i) = F_i(d; \hat{\boldsymbol{\theta}}_i), \tag{2.71}$$

and the performance, assuming model m_j is true, is given by

$$h_j = P(Y \leq d \mid m_j) = F_j(d; \hat{\boldsymbol{\theta}}_j), \tag{2.72}$$

where, because $\mathcal{A} = \mathcal{M}'$, we have $h_j = v_j$, $j = 1, 2$. Assuming model m_j is true, model m_i gives a conservative prediction of $P(Y \leq d)$ if $F_i(d; \hat{\boldsymbol{\theta}}_i) \leq F_j(d; \hat{\boldsymbol{\theta}}_j)$, and

a non-conservative prediction otherwise. Accordingly, we use the following penalty function

$$\psi(m_i, m_j) = \tilde{\psi}(h_i, h_j) = \begin{cases} \beta_1(h_i - h_j)^2 & \text{if } h_i \leq h_j, \text{ and} \\ \beta_2(h_i - h_j)^2 & \text{if } h_i > h_j. \end{cases} \tag{2.73}$$

To illustrate, consider the identical $n = 5$ samples from Example 2.3, and let parameters $d = 6$, $\beta_1 = 100$ and $\beta_2 = 1000$. By Eq. (2.71), $h_1 = 0.978$ and $h_2 = 0.963$, so that, by Eq. (2.68), model m_2 is conservative with respect to m_1. Using $p_1 = \bar{l}_1 = 0.453$, $p_2 = \bar{l}_2 = 0.547$, by Eq. (2.45), we have

$$\begin{aligned} u(m_1) &= 0 \cdot p_1 + \beta_2(h_1 - h_2)^2 p_2 = 0.123 \\ u(m_2) &= \beta_1(h_2 - h_1)^2 p_1 + 0 \cdot p_2 = 0.0102 \end{aligned} \tag{2.74}$$

so that model m_2 is optimal. $\Diamond$

2.4 Sensitivity analysis

As discussed in Section 2.1.2.4, $m^\star \in \mathcal{M}$ depends on the: (1) available information, $\mathcal{I}$, (2) class of candidate models, $\mathcal{M}$, and in the case of the decision-theoretic method for model selection, (3) model use. Changes in any or all of these items may result in a different $m^\star$. To illustrate, we consider the following example.

Example 2.10: Recall Example 2.8, where the objective was to select the optimal model for random variable Z that was consistent with the model use (design).

- $\mathcal{M}_1 = \{F_1(z; \hat{\boldsymbol{\theta}}_1), F_2(z; \hat{\boldsymbol{\theta}}_2)\}$, where F_1 and F_2 denote the CDFs for a Gaussian and lognormal random variable, respectively.
- $\bar{l}_1 = 0.453$ and $\bar{l}_2 = 0.547$ with $n = 5$ samples of Z.
- The expected utilities were $u(a_1) = 2.67$ and $u(a_2) = .594$; parameters $p_1 = \bar{l}_1$, $p_2 = \bar{l}_2$, $\beta_1 = 100$, $\beta_2 = 1000$, and $\gamma_0 = 1$ were used for calculations.

Under both the method of maximum likelihood and the decision-theoretic method, model $m^\star = m_2$. We next consider three cases, corresponding to changes in $\mathcal{M}_1$, $\mathcal{I}$, and model use, respectively. **Case #1**: suppose we add model $F_3(z; \hat{\boldsymbol{\theta}}_3)$ to the collection of competing models, where F_3 is the CDF for an exponential random variable, *i.e.*, Eq. (2.16), and $\hat{\boldsymbol{\theta}}_3$ is the maximum likelihood estimate for $\boldsymbol{\theta}_3 = (\alpha, \nu)^T$. The new class of candidate models is a superset of the original class, *i.e.*,

$$\mathcal{M}_1 \subset \mathcal{M}_2 = \{F_1(z; \hat{\boldsymbol{\theta}}_1), F_2(z; \hat{\boldsymbol{\theta}}_2), F_3(z; \hat{\boldsymbol{\theta}}_3)\}. \tag{2.75}$$

Table 2.1. Sensitivity of optimal model.

	Case #1			Case #2			Case #3		
	m_1	m_2	m_3	m_1	m_2	m_3	m_1	m_2	m_3
$\bar{l}_i$	0.188	0.228	**0.584**	**0.473**	0.386	0.141	**0.473**	0.386	0.141
$u(a_i)$	115	44.0	**4.85**	200	78.3	**17.2**	30.1	**12.8**	17.2

Case #2: suppose that, in addition to Case #1, more information on Z becomes available, *e.g.*, we have $n + n'$ total samples for Z, where $n' \geq 1$. **Case #3**: suppose that, in addition to Cases #1 and #2, the model use is modified, *e.g.*, parameter β_2 is replaced by $\beta_2' \neq \beta_2$.

The likelihoods and expected utilities for each case are shown in Table 2.1. When the collection of candidate models is augmented by one (Case #1), both the method of maximum likelihood and the decision-theoretic method select the newest model, m_3. Under increasing data (Case #2), the method of maximum likelihood selects a different model, m_1, than in Case #1; parameter $n' = 5$ was used for calculations. The expected utilities also change under the added data but, under the decision-theoretic method, the optimal model remains the same as in Case #1. For Case #3, let $\beta_2' = 150$, meaning that we assign less penalty for non-conservative designs than with $\beta_2 = 1000$. As shown in Table 2.1, a different model, m_2, is selected under the decision-theoretic approach; the optimal model under the method of maximum likelihood is unchanged from Case #2 since this method for model selection is independent of model use. $\Diamond$

Chapter 3

Applications of model selection

Consider, as in Section 2.1.1, the input/output relationship given by

$$Y = \mathcal{D}(Z; X), \tag{3.1}$$

where Z and Y are real-valued stochastic processes, respectively, and $\mathcal{D}$ is defined by a functional form and a single coefficient, denoted by X. Information on Z, X, and the functional form for $\mathcal{D}$ is limited. Three cases are considered. First, in Section 3.1 we assume complete information on input Z and functional form for $\mathcal{D}$, and apply methods for model selection to choose optimal models for coefficient X. Second, in Section 3.2 we assume complete information on X and $\mathcal{D}$, and apply methods for model selection to choose optimal models for Z. The design of a cantilever beam is used for illustration in Sections 3.1 and 3.2. Third, in Section 3.3 we assume complete information on X and Z, and apply methods for model selection to choose optimal models for the functional form for $\mathcal{D}$; the performance of a two degree-of-freedom oscillator is used for illustration. The organization of the ensuing discussion is further illustrated by Table 3.1; more general combinations of complete and incomplete information are possible.

Table 3.1. Organization of Chapter 3.

Section	Available information		
	X	Z	$\mathcal{D}$
3.1	Incomplete	Complete	Complete
3.2	Complete	Incomplete	Complete
3.3	Complete	Complete	Incomplete

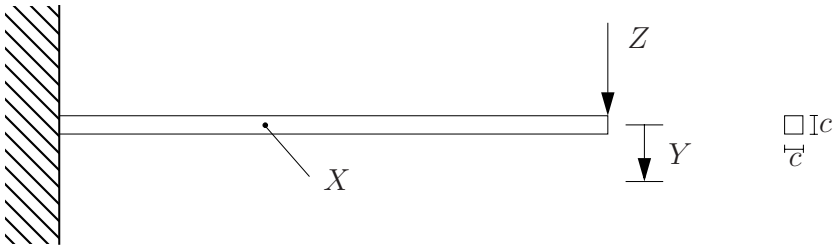

Figure 3.1: Design of cantilever beam.

3.1 Incomplete information on random variable

Consider the cantilever beam shown in Fig. 3.1, having fixed length and square cross section of dimension, $c > 0$. The flexibility of the beam is denoted by X, the applied tip load is denoted by Z, and the tip deflection of the beam is denoted by Y; c is a deterministic design variable. The information on X is incomplete. We assume: (1) X can be modeled as a random variable with unknown probability law, (2) Z is perfectly known and equal to one, (3) the functional form for $\mathcal{D}$ is given by

$$Y = \mathcal{D}(Z; X) = \frac{1}{c^4} X, \tag{3.2}$$

which is consistent with Euler-Bernoulli beam theory [56], and (4) the design constraint for the beam is given by

$$P\left(g(Y) \leq d\right) = \bar{q}, \tag{3.3}$$

where $g(Y) = Y$ is the property of interest, and $d \geq 0$ and $\bar{q} \in [0, 1]$ are the maximum tip displacement and required reliability of the design, respectively. By Eq. (3.3), we note that the model use is design. The objectives of this study are: (i) to select optimal models for the random flexibility, X, and (ii) to design the beam cross section, c, to satisfy Eq. (3.3).

3.1.1 Available information

It is assumed that the information on X is limited to n independent samples, denoted by $\{x_1, x_2, \ldots, x_n\}$. In addition, because X is a flexibility, prior knowledge requires that X must be real and nonnegative almost surely, *i.e.*,

$$P(X \geq 0) = 1. \tag{3.4}$$

3.1.2 Candidate models

Consider the following collection of candidate models for X,

$$\mathcal{M}' = \{m_1, m_2, m_3\} = \{F_1(x; \hat{\boldsymbol{\theta}}_1), F_2(x; \hat{\boldsymbol{\theta}}_2), F_3(x; \hat{\boldsymbol{\theta}}_3)\}, \tag{3.5}$$

where we have used $\mathcal{M}'$ to denote that this is a finite subcollection of $\mathcal{M}$, the class of all candidate models for X that is consistent with the available information. The elements of $\mathcal{M}'$ are given by

$$\begin{aligned} F_1(x; \boldsymbol{\theta}_1) &= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{s-\mu}{\sigma}\right)^2\right] \mathrm{d}s, \\ F_2(x; \boldsymbol{\theta}_2) &= \int_{0}^{x} \frac{1}{\sqrt{2\pi}\zeta s} \exp\left[-\frac{1}{2}\left(\frac{\ln s-\eta}{\zeta}\right)^2\right] \mathrm{d}s, \text{ and} \\ F_3(x; \boldsymbol{\theta}_3) &= \int_{\alpha}^{x} \nu \exp\left[-\nu(s-\alpha)\right] \mathrm{d}s. \end{aligned} \tag{3.6}$$

where $\boldsymbol{\theta}_1 = (\mu, \sigma)^T$, $\boldsymbol{\theta}_2 = (\eta, \zeta)^T$, and $\boldsymbol{\theta}_3 = (\alpha, \nu)^T$ denote the model parameters. By Eq. (3.6), F_1 is the probability law for a Gaussian random variable, F_2 is the probability law for a lognormal random variable, and F_3 is the probability law for an exponential random variable. The Gaussian model, m_1, is commonly used and will be included in this study. However, because it violates Eq. (3.4), it is inconsistent with the available information.

For calculations, the samples, $\{x_i\}$, $i = 1, 2, \ldots, n$, are generated from a lognormal random variable with $\mathrm{E}[X] = 5$, $\mathrm{Var}[X] = 2$. Because of this, we refer to m_2 as the "true model". Multiple data sets of different lengths are to be considered. For each candidate model, m_i, and each data set, the parameter vector $\boldsymbol{\theta}_i$, is a maximum likelihood point estimate, $\hat{\boldsymbol{\theta}}_i$ (see Section 2.2.1). Figure 3.2 illustrates the PDFs of the three models, denoted by $f_i(x; \hat{\boldsymbol{\theta}}_i)$, $i = 1, 2, 3$, for $n = 5$ and two different data sets; the data are also shown. Note that the models for X are sensitive to the particular data set used, as indicated by the differences between the plots on the left and right in Fig. 3.2.

If $m_i \in \mathcal{M}'$ is used as the model for X, we assume the design of the beam cross section to satisfy

$$F_i\left(c_i^4 d; \hat{\boldsymbol{\theta}}_i\right) = \bar{q}. \tag{3.7}$$

By Eqs. (3.3) and (3.7) and the behavior of the right-hand tails of the models, it can be shown that m_3 always yields a conservative design of the beam, while m_1 a non-conservative design.

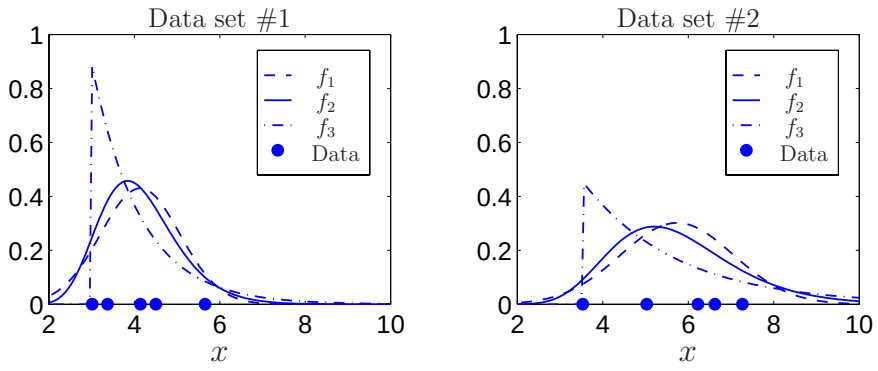

Figure 3.2: Candidate models for flexibility, X, when $n = 5$.

3.1.3 Model selection

The method of maximum likelihood and the decision-theoretic method for model selection developed in Chapter 2 are applied to: (1) select a model for X, and (2) select a value for the design parameter, c. Different models for X and values for c will be selected under the two methods for analysis; only the decision-theoretic method accounts for model use and, hence, the consequences of selecting an inappropriate model for X.

3.1.3.1 Method of maximum likelihood

Table 3.2 lists the normalized likelihoods, $\bar{l}_i$, $i = 1, 2, 3$, of each of the three models in $\mathcal{M}'$ under ten different data sets, each of size $n = 5$. As in Section 3.1.2, all data were sampled from a lognormal random variable. For each data set, the likelihood of the optimal model is in bold. Results show that if n is small, the optimal model depends strongly on the particular data set used for calculations. Model m_1, which is inconsistent with the available information, is selected three times. Further, the true model, m_2, is not chosen for all ten data sets.

The results in Table 3.2 are for limited information ($n = 5$). Figure 3.3 shows how increasing sample size affects the model selection process for four of the ten data sets from Table 3.2. In each plot, the likelihood of the true model is a solid line; the likelihoods of the others are shown with dashed lines. For small n, any candidate model in $\mathcal{M}'$ can be chosen. The true model is selected for data sets #2, #5, and #8 when $n > 30$, but the inconsistent model, m_1, is selected for data set #1 when $n = 100$. This illustrates that even with a large sample, an inappropriate model can be selected. As n increases beyond 100, m_2 is selected for all data sets considered. However, for finite n, this method of model selection does not guarantee that the design will satisfy the performance requirement of Eq. (3.3).

Table 3.2. Normalized likelihoods, $\bar{l}_i$, of models for X with $n = 5$.

Data set	m_1	m_2	m_3
1	0.1814	0.2317	**0.5869**
2	**0.4206**	0.2988	0.2806
3	**0.4157**	0.3671	0.2172
4	0.2273	0.2722	**0.5006**
5	0.2374	0.2164	**0.5462**
6	0.0876	0.1353	**0.7771**
7	0.1392	0.1826	**0.6782**
8	**0.4269**	0.3143	0.2588
9	0.3093	0.3062	**0.3844**
10	0.2102	0.2442	**0.5456**

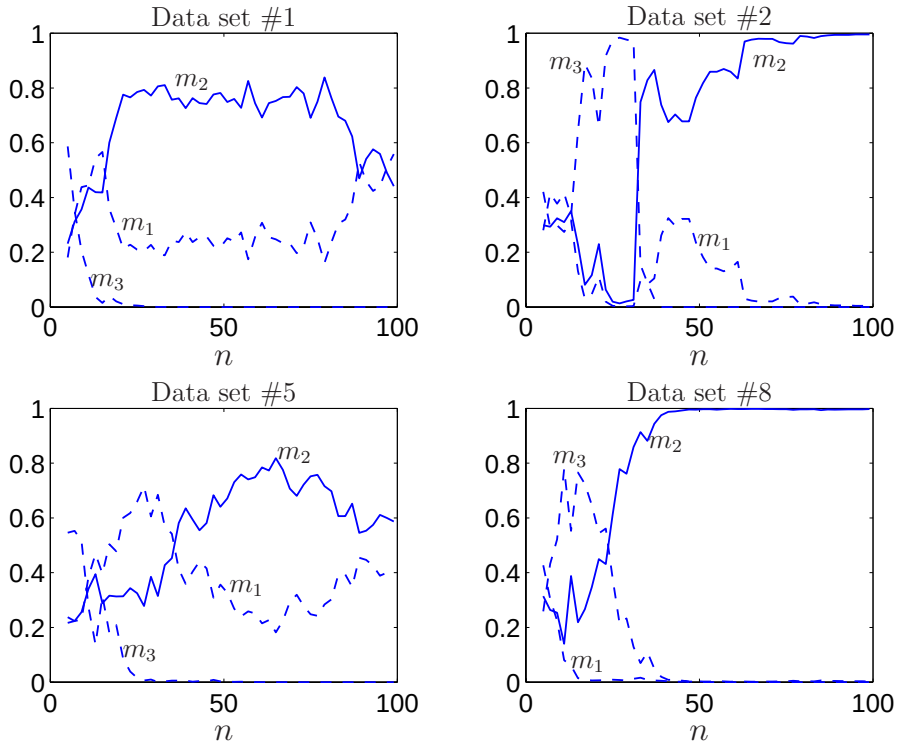

Figure 3.3: Normalized likelihoods, $\bar{l}_i$, of models for X *vs.* sample size, n.

3.1.3.2 Decision-theoretic method

The decision-theoretic method for model selection is now applied to the beam design problem. The collection of possible models for X is given by Eq. (3.5); the collection of candidate designs is given by

$$\mathcal{A} = \{a_1, a_2, a_3\} = \{c_1, c_2, c_3\}, \tag{3.8}$$

where c_i, given by Eq. (3.7), is the value for the beam cross section under model $m_i \in \mathcal{M}'$, $i = 1, 2, 3$.

For design a_i under model m_j, we have

$$P(g(Y) \leq d \mid a_i, m_j) = F_j(c_i^4 d; \hat{\boldsymbol{\theta}}_j). \tag{3.9}$$

By Eqs. (3.7) and (3.9), the design constraint, Eq. (3.3), is satisfied if $i = j$ and violated otherwise. Accordingly, define ξ_{ij} such that

$$F_j(c_i^4 \xi_{ij}; \hat{\boldsymbol{\theta}}_j) = \bar{q}. \tag{3.10}$$

It follows that

$$\xi_{ij} \leq d \iff P(g(Y) \leq d \mid a_i, m_j) \geq \bar{q}, \text{ and} \tag{3.11a}$$
$$\xi_{ij} > d \iff P(g(Y) \leq d \mid a_i, m_j) < \bar{q}. \tag{3.11b}$$

If a_i satisfies Eq. (3.11a), it is a conservative design for the beam, while if a_i satisfies Eq. (3.11b), it is a non-conservative design of the beam.

Recall that the utility function, Eq. (2.44), has two parts: (1) the cost function, γ, and (2) the penalty function, ψ. Because the model use is design (see Eq. (3.3)), the utility formulation of Section 2.3.2.2 applies. Let

$$\gamma(c_i) = \gamma_0 c_i^2, \; i = 1, 2, 3, \tag{3.12}$$

where $\gamma_0 > 0$ is a deterministic scaling parameter, define the cost of design a_i. By Eq. (3.12), a model for X that requires a large beam cross section has a high cost. An appropriate penalty function is given by

$$\psi(a_i, m_j) = \tilde{\psi}(\xi_{ij}, d) = \begin{cases} \beta_1(\xi_{ij} - d)^2 & \text{if } \xi_{ij} \leq d, \text{ and} \\ \beta_2(\xi_{ij} - d)^2 & \text{if } \xi_{ij} > d, \end{cases} \tag{3.13}$$

where $\beta_2 > \beta_1 \geq 0$ are deterministic parameters; ψ is identical to Eq. (2.53) discussed in Section 2.3.2.2. By Eq. (3.13), a non-conservative design is heavily penalized, as is a design that is overly conservative.

The expected utility of each design is given by Eq. (2.45), where $p_i = \bar{l}_i$ is the probability that m_i is true, $i = 1, 2, 3$; the optimal design minimizes Eq. (2.45).

Table 3.3. Expected utilities, $u(a_i)$, with $n = 5$.

Data set	a_1	a_2	a_3
1	2,299	1,124	**1.239**
2	3,792	1,426	**2.481**
3	1,074	814.6	**1.648**
4	4,041	1,213	**1.618**
5	4,397	783.5	**1.644**
6	1,197	499.9	**1.075**
7	2,025	804.8	**1.183**
8	2,747	1,365	**2.301**
9	3,280	1,501	**1.853**
10	2,659	1,199	**1.449**

Shown in Table 3.3 are the expected utilities of each design for the identical 10 data sets, each of size $n = 5$, that were used in Table 3.2. By Eq. (3.7), there is a one-to-one mapping between the members of $\mathcal{M}'$ and the members of $\mathcal{A}$. Hence if $a_i \in \mathcal{A}$ is the optimal design, we say that $m_i \in \mathcal{M}'$ is the optimal model for X. This allows a direct comparison between the results using the method of maximum likelihood, discussed in Section 3.1.3.1, and the results using the decision-theoretic method. With such a small data set, one cannot hope to select the true model, m_2. However, unlike the results of Section 3.1.3.1, the optimal model is insensitive to the particular data set used; model m_3 is selected every time, and the non-conservative model, m_1, was never chosen. Parameters $d = 9.2$, $\bar{q} = 0.99$, $\gamma_0 = 1$, $\beta_1 = 1/5$, and $\beta_2 = 500$ were used for calculations.

Shown in Fig. 3.4 are the expected utilities for the data sets from Fig. 3.3. In each case, the expected utility corresponding to the true model is plotted with a solid line; the expected utility corresponding to the others is shown with dashed lines. As illustrated, the selection process is less sensitive to data set and sample size than the results in Fig. 3.3. As n increases, the true model is selected for any data set. Further, when an incorrect model is chosen it is consistently m_3, which is conservative as defined by Eq. (3.11a). This conservatism is further illustrated by Fig. 3.5, which shows the value for the design parameter, c, under the optimal model using the method of maximum likelihood, c_{ML}, and the decision-theoretic approach, c_{DT}. Note that $c_{\mathrm{DT}} \geq c_{\mathrm{ML}}$ for all data sets and all n because a conservative beam design has a larger cross section. As $n \to \infty$, c_{DT} and c_{ML} coincide.

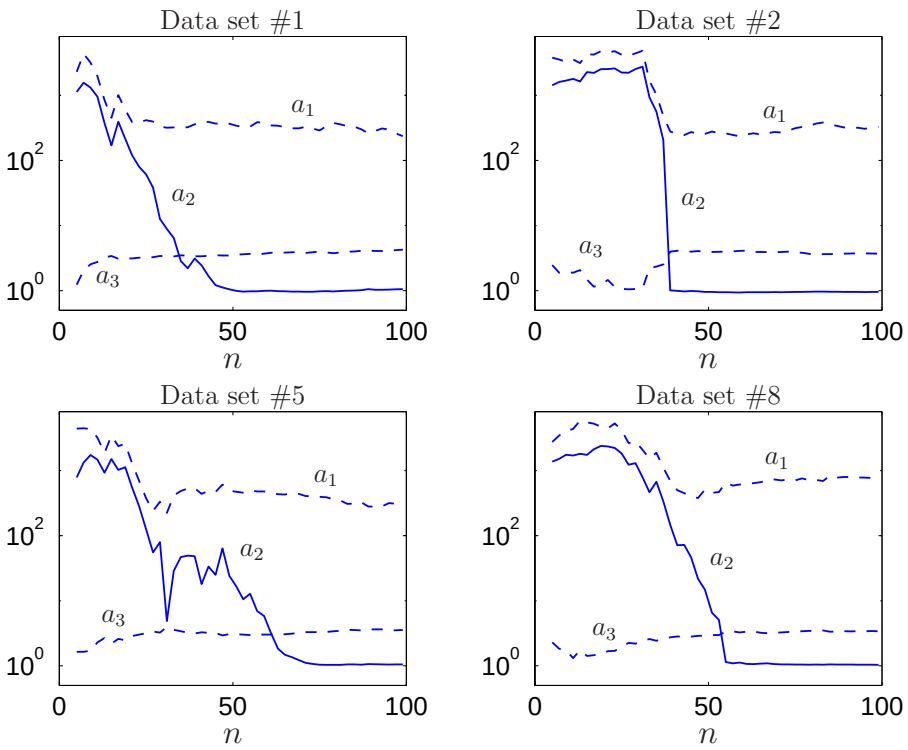

Figure 3.4: Expected utilities, $u(a_i)$, *vs.* sample size, n.

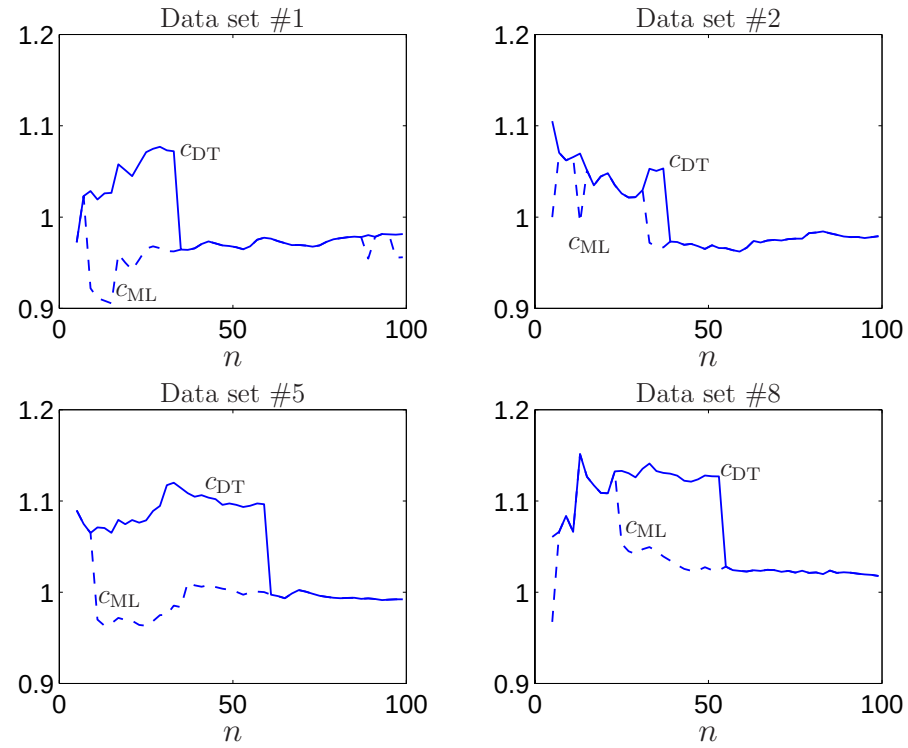

Figure 3.5: Beam cross section, c, under both methods for model selection.

3.2 Incomplete information on stochastic process

Consider again the cantilever beam shown in Fig. 3.1, where now the information on Z is incomplete. We assume: (1) Z is time-varying and can be modeled as a stochastic process with unknown probability law, (2) X is perfectly known, deterministic, and equal to one, (3) the functional form for $\mathcal{D}$ is perfectly known, and (4) the design constraint for the beam is given by Eq. (3.3), where

$$g(Y) = \max_{t \in T} \mid Y(t) \mid = \frac{1}{c^4} \max_{t \in T} \mid Z(t) \mid \tag{3.14}$$

is the property of Y of interest, and T is an arbitrary index set. The objectives of this study are: (i) to select optimal models for the stochastic tip load, Z, and (ii) to design the beam cross section, c, to satisfy Eq. (3.3). We consider discrete-time and continuous-time models for Z in Sections 3.2.1 and 3.2.2, respectively.

3.2.1 Discrete-time models

We first consider a class of discrete-time models for Z, so that index set $T = \{1, 2, \ldots, n\}$ is finite. Let

$$Z_i(t) = \theta_{i,0} + \sum_{j=1}^{i} \theta_{i,j} Z_i(t-j) + W_i(t), \ t = 1, 2, \ldots, n, \tag{3.15}$$

where $\boldsymbol{\theta}_i = (\theta_{i,0}, \theta_{i,1}, \ldots, \theta_{i,i})^T$ denotes a vector of deterministic parameters, and $W_i(t) \sim$ iid $N(0, 1/\zeta_i)$ is zero mean, Gaussian white noise with variance $1/\zeta_i$. By Eq. (3.15), $Z_i(t)$ is an autoregressive stochastic process of order i [16], abbreviated as AR(i). Models of this form are also referred to as **time series** models. Vector $\mathbf{Z}_0 = \{Z_i(0), Z_i(-1), \ldots, Z_i(1-i)\}$ are the initial conditions to the model. Process Z_i is stationary if $\mathbf{Z}_0 \sim N(\mathbf{0}, \mathbf{c}_i)$, where $\mathbf{c}_i$ denotes the stationary covariance function of an AR(i) process.

3.2.1.1 Available information

We assume two samples of $Z(t)$, denoted by $\{z_k(1), \ldots, z_k(n)\}$, $k = 1, 2$, are available; the collection $\{z_k(1), \ldots, z_k(n)\}$, $k = 1, 2$, defines a data set. For calculations, the data set is taken from a stationary AR(3) process with $\boldsymbol{\theta}_3 = (0, 0.7, -0.5, -0.3)^T$, $\zeta_3 = 0.65$. To enforce stationarity, the data set was generated with initial conditions $\mathbf{Z}_0 \sim N(\mathbf{0}, \mathbf{c}_3)$. However, the only knowledge of Z comes from the data set; the parameter vector and order of Eq. (3.15) are unknown.

Prior knowledge on the collection of candidate models for Z is quantified by: (1) r, the order of the largest model considered in the analysis, (2) $f'(\boldsymbol{\theta}_i, \zeta_i)$, the

joint prior PDFs for the unknown parameters of each candidate model in $\mathcal{M}'$, and (3) p'_i, the prior probability that each $m_i \in \mathcal{M}'$ is true. For calculations, $r = 4$ and non-informative prior probabilities are used, *i.e.*, $p'_i = 1/r$, $i = 1, 2, 3, 4$. Any further knowledge on the physics of Z can be applied through appropriate choice of $f'(\boldsymbol{\theta}_i, \zeta_i)$.

3.2.1.2 Candidate models

Consider the following collection of candidate models for Z,

$$\begin{aligned}\mathcal{M}' &= \{m_1, m_2, m_3, m_4\} \\ &= \{Z_1(t; \boldsymbol{\theta}_1, \zeta_1), Z_2(t; \boldsymbol{\theta}_2, \zeta_2), Z_3(t; \boldsymbol{\theta}_3, \zeta_3), Z_4(t; \boldsymbol{\theta}_4, \zeta_4)\},\end{aligned} \tag{3.16}$$

where $\mathcal{M}'$ is used to denote that the class of models defined by Eq. (3.16) is a finite subcollection of $\mathcal{M}$, the class of all models for Z that are consistent with the available information. By Eq. (3.15), model $m_i \in \mathcal{M}'$ is an AR(i) process; we refer to m_3 as the "true model" because the available information was generated by an AR(3) process. The data set, $\{z_k(1), \ldots, z_k(n)\}$, $k = 1, 2$, is not known to come from a stationary source. Hence, in the calculations that follow, the initial conditions, $\mathbf{Z}_0$, are set to zero for each model in $\mathcal{M}'$.

We calculate

$$g(Y) \mid m_i = \max_{t \in [1,n]} \mid Z_i(t) \mid, \; i = 1, 2, 3, 4, \tag{3.17}$$

which gives estimates of Eq. (3.14) under each candidate model. If $m_i \in \mathcal{M}'$ is used as the model for Z, we assume the design of the beam cross section to be given by Eq. (3.7) where F_i denotes the CDF of $g(Y)|m_i$.

3.2.1.3 Model selection

Bayes' method and the decision-theoretic method for model selection developed in Chapter 2 are applied to: (1) select a model for Z, and (2) select a value for the design parameter, c. Different models for Z and values for c will be selected under the two methods for analysis.

3.2.1.3.1 Bayes' method The prior joint PDF of $(\boldsymbol{\theta}_i, \zeta_i)$ is assumed to have a Normal-Gamma distribution [81], pp. 226-227, with parameters $(\boldsymbol{\phi}'_i, \boldsymbol{\sigma}'_i, \eta'_i, \nu'_i)$, *i.e.*,

$$\boldsymbol{\theta}_i \mid \zeta_i \sim N(\boldsymbol{\phi}'_i, \boldsymbol{\sigma}'_i \mid \zeta_i), \tag{3.18}$$

where N denotes the probability law for a Gaussian random vector, and

$$f'_{\zeta_i}(y; \eta'_i, \nu'_i) = \frac{(\eta'_i \nu'_i)^{\nu'_i/2}}{\Gamma(\nu'_i/2)} y^{\nu'_i/2 - 1} \exp\left(-\frac{\eta'_i \nu'_i y}{2}\right) \tag{3.19}$$

is the prior PDF for ζ_i. This is a conjugate prior density (see [81], pp. 318-319); assuming $\boldsymbol{\sigma}_i'$ has full rank, $i = 1, 2, \ldots, r$, the posterior density is also Normal-Gamma with parameters $(\boldsymbol{\phi}_i'', \boldsymbol{\sigma}_i'', \eta_i'', \nu_i'')$, where $\mathbf{z}_i = (z(i+1), \ldots, z(n))^T$,

$$\begin{aligned}
(\boldsymbol{\sigma}_i'')^{-1} &= (\boldsymbol{\sigma}_i')^{-1} + \mathbf{v}_i^T \mathbf{v}_i, \\
\boldsymbol{\phi}_i'' &= \boldsymbol{\sigma}_i'' \left[(\boldsymbol{\sigma}_i')^{-1} \boldsymbol{\phi}_i' + \mathbf{v}_i^T \mathbf{z}_i \right], \\
\eta_i'' \nu_i'' &= \eta_i' \nu_i' + (\boldsymbol{\phi}_i')^T (\boldsymbol{\sigma}_i')^{-1} \boldsymbol{\phi}_i' - (\boldsymbol{\phi}_i'')^T (\boldsymbol{\sigma}_i'')^{-1} \boldsymbol{\phi}_i'' + \mathbf{z}_i^T \mathbf{z}_i, \text{ and} \\
\nu_i'' &= \nu_i' + n - i,
\end{aligned} \tag{3.20}$$

and

$$\mathbf{v}_i = \begin{bmatrix} 1 & z(i) & \ldots & z(1) \\ 1 & z(i+1) & \ldots & z(2) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & z(n-1) & \ldots & z(n-i) \end{bmatrix} \tag{3.21}$$

Estimates $\hat{\boldsymbol{\theta}}_i = \mathrm{E}[\boldsymbol{\theta}_i]$ and $\hat{\zeta}_i = \mathrm{E}[\zeta_i]$ are then used to create samples of $Z_i(t)$, $i = 1, 2, 3, 4$, via Eq. (3.15). The posterior probability that $m_i \in \mathcal{M}'$ is true is given by (see [50], pp. 752-753)

$$p_i'' = \kappa p_i' f_i(\mathbf{z}_i), \ i = 1, 2, \ldots, r, \tag{3.22}$$

where

$$\kappa^{-1} = \sum_{j=1}^{r} p_j' f_j(\mathbf{z}_j), \tag{3.23}$$

$\Gamma(\,\cdot\,)$ denotes the gamma function, and

$$f_i(\mathbf{z}_i) = (2\pi)^{-(\nu_i'' - \nu_i')/2} \frac{|\,\boldsymbol{\sigma}_i''\,|^{1/2}}{|\,\boldsymbol{\sigma}_i'\,|^{1/2}} \frac{(\eta_i' \nu_i'/2)^{\nu_i'/2}}{(\eta_i'' \nu_i''/2)^{\nu_i''/2}} \frac{\Gamma(\nu_i''/2)}{\Gamma(\nu_i'/2)} \tag{3.24}$$

Table 3.4 lists the posterior probabilities, p_i'', $i = 1, 2, 3, 4$, of each of the four models in $\mathcal{M}'$ with $n = 34$; the probability of the optimal model is in bold. The probabilities are tabulated for five independent data sets, *i.e.*, five independent collections of $\{z_k(1), \ldots, z_k(n)\}$, $k = 1, 2$. Results indicate that the selection process is again sensitive to the particular data set used. As in Section 3.1.3.1, under Bayes' method for model selection, there is no guarantee that the design will satisfy the performance requirement of Eq. (3.3).

This sensitivity to data set is high when data is limited, but decreases for increasing sample size, n. This is illustrated by Fig. 3.6 for four of the five data sets. In each plot, the posterior probability of the true model, m_3, is a solid line; the posterior probabilities of the others are shown with dashed lines. In data sets #1, #2, and #4, there is no clear winner among models m_2, m_3, and m_4 for $n < 50$. Model m_1 is inappropriate in all cases for any n. As n increases, the true model is selected in every case. The AR(4) model is chosen for data set #2; it can be shown that this is a degenerate AR(3) model because, as n grows large, $\hat{\theta}_4 = \mathrm{E}[\theta_4]$ approaches zero.

Table 3.4. Posterior probabilities, p_i'', of models for Z with $n = 34$.

Data set	m_1	m_2	m_3	m_4
1	5.075×10^{-3}	**0.5497**	0.3987	4.654×10^{-2}
2	3.036×10^{-6}	1.030×10^{-2}	**0.6040**	0.3857
3	2.468×10^{-5}	0.2357	**0.7620**	2.248×10^{-3}
4	3.132×10^{-6}	**0.5526**	0.4470	3.371×10^{-4}
5	1.010×10^{-4}	2.641×10^{-2}	**0.8632**	0.1103

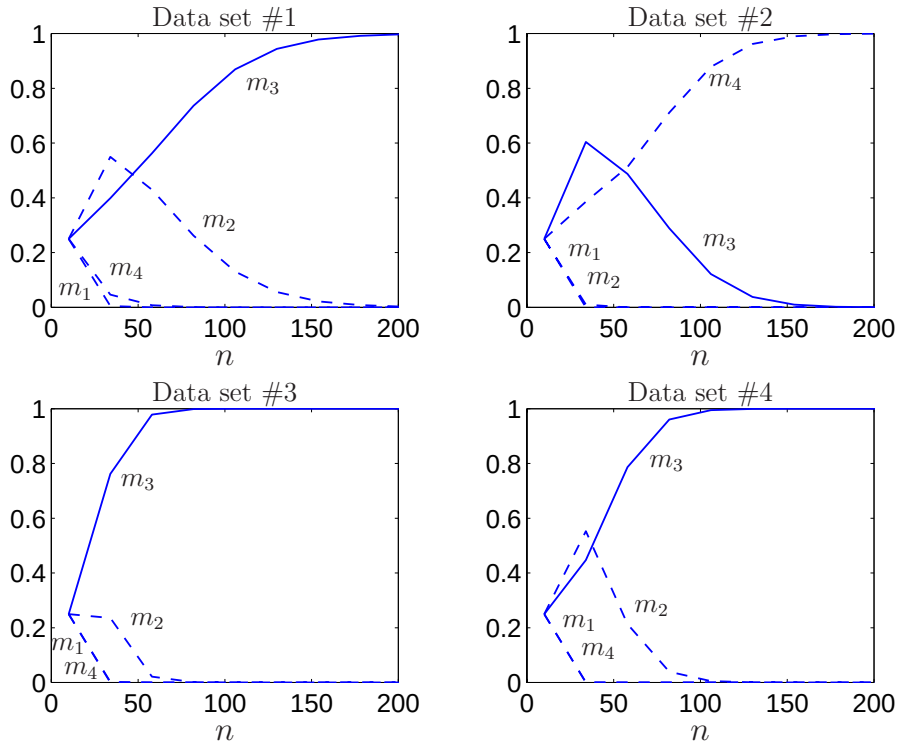

Figure 3.6: Posterior probabilities, p_i'', of models for Z *vs.* sample size, n.

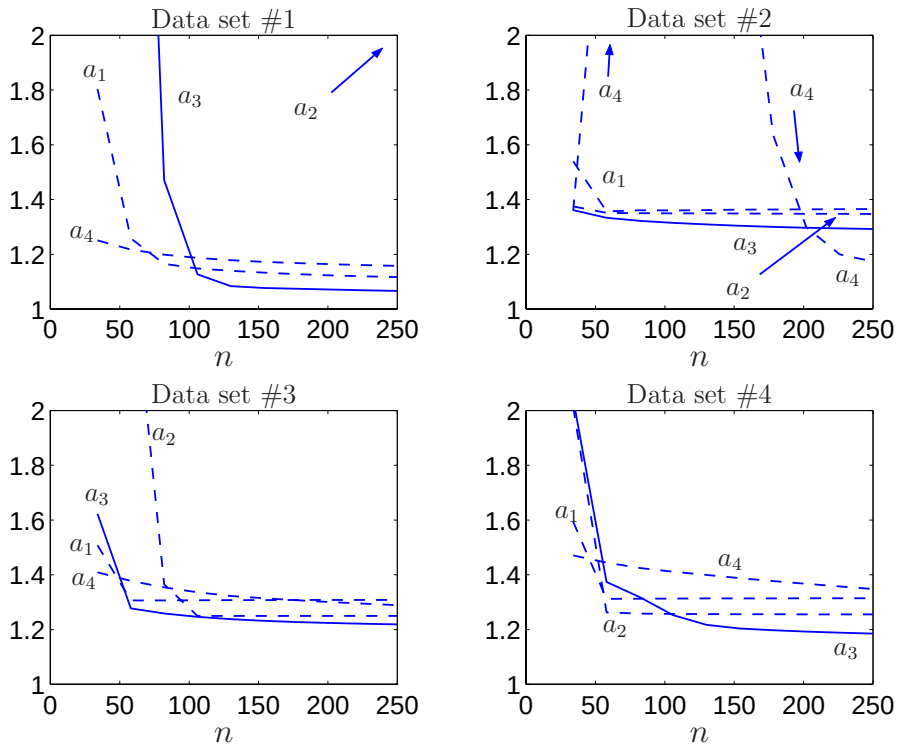

Figure 3.7: Expected utilities, $u(a_i)$, *vs.* sample size, n.

3.2.1.3.2 Decision-theoretic method We next apply the decision-theoretic method for model selection to choose the optimal design of the beam. The derivation of the utility function is identical to that of Section 3.1.3.2, with X everywhere replaced by $g(Y)|m_i = \max_{t\in[1,n]} |Z_i(t)|$; wherever the CDF of $g(Y)|m_i$ is required, *e.g.*, Eq. (3.10), it is estimated via Monte Carlo simulation using 500 samples. The expected utility of each design is computed by Eq. (2.45), with $p_i = p_i''$.

Shown in Fig. 3.7 are the expected utilities of each design for the data sets from Fig. 3.6. As before, by Eq. (3.7) there is a one-to-one mapping between the members of $\mathcal{M}'$ and the members of $\mathcal{A}$. Hence if $a_i \in \mathcal{A}$ is the optimal design, we say that $m_i \in \mathcal{M}'$ is the optimal model for X. For each data set shown in Fig. 3.7, the expected utility corresponding to the true model is shown with a solid line, while the expected utility corresponding to the others is shown with dashed lines. The optimal model changes from data set to data set and for different values of n. This is due to the fact that, contrary to the class of candidate models for X defined by Eq. (3.5), here all members of $\mathcal{M}'$ are consistent with the physics of the problem and have the same functional form. The same behavior is still observed: for small n, a conservative model is selected; as n grows, the true model is eventually selected.

Because of the form for the utility function, the decision-theoretic method for model selection will favor a conservative design of the beam. This is illustrated by Fig. 3.8, which shows the value for the design parameter, c, under the optimal model using Bayes' method, denoted by c_{B}, and the decision-theoretic approach, denoted

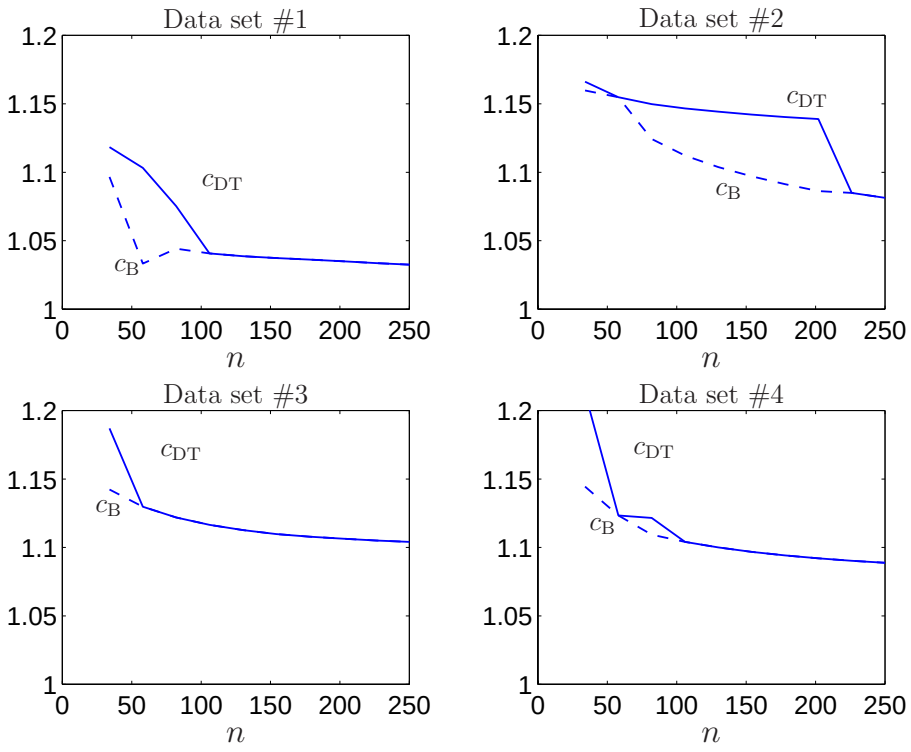

Figure 3.8: Beam cross section, c, under both methods for model selection.

by c_{DT}. Note that $c_{\mathrm{DT}} \geq c_{\mathrm{B}}$ for all data sets and all n because a conservative design has a larger cross section; as $n \to \infty$, c_{DT} and c_{B} coincide.

3.2.1.4 Model use

In this section we emphasize that, under the decision-theoretic method for model selection, a model is optimal for a given model use; if the model use changes, a different model may be selected. To illustrate consider, for $t = 1, 2, \ldots, n$, the AR(1) and AR(2) time series models for Z defined by Eq. (3.15), *i.e.*,

$$\begin{aligned} m_1\colon&\ Z_1(t) = \theta_{1,0} + \theta_{1,1} Z_1(t-1) + W_1(t), \\ m_2\colon&\ Z_2(t) = \theta_{2,0} + \theta_{2,1} Z_2(t-1) + \theta_{2,2} Z_2(t-2) + W_2(t), \end{aligned} \tag{3.25}$$

where W_1 and W_2 are zero-mean, Gaussian white noise with variances $1/\zeta_1$ and $1/\zeta_2$, respectively. Suppose further we are interested in the following two design constraints on output Y

$$\begin{aligned} &P\left(g_1(Y) \leq d_1\right) = \bar{q}, \text{ and} \\ &P\left(g_2(Y) \leq d_2\right) = \bar{q}, \end{aligned} \tag{3.26}$$

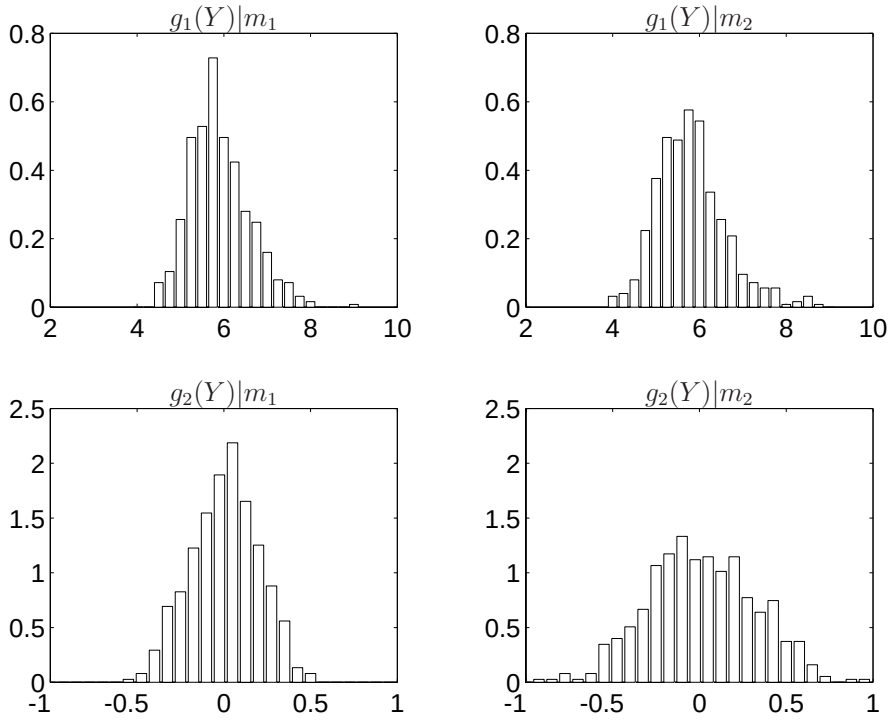

Figure 3.9: Histograms of extremes (top) and sums (bottom) of Y.

where d_1, d_2, and $\bar{q} \in [0, 1]$ are deterministic parameters, and

$$g_1(Y) = \max_{t \in [1,n]} \mid Y(t) \mid, \text{ and}$$
$$g_2(Y) = \frac{1}{n} \sum_{t=1}^{n} Y(t) \tag{3.27}$$

define two properties of Y. Property $g_1(Y)$ is based on the extremes of Y and is identical to Eq. (3.14); property $g_2(Y)$ is based on the sum of Y and may be used, for example, to quantify damage accumulation in a structural system.

Histograms of $g_1(Y)|m_i$, $i = 1, 2$, are shown in the top two plots of Fig. 3.9, and histograms of $g_2(Y)|m_i$, $i = 1, 2$, are shown in the bottom two plots of Fig. 3.9; the results are estimates using 500 Monte Carlo simulations. In all cases, we assume $c = 1$ so that, by Eq. (3.14), $\max_{t \in T} |Y(t)| = \max_{t \in T} |Z(t)|$. It is evident that while the distribution of the extremes are similar, the distribution of the sums are not. Parameters $\boldsymbol{\theta}_1 = (0, 0.8201)^T$, $\boldsymbol{\theta}_2 = (0, 0.5, 0.4)^T$, $1/\zeta_1 = 1.223$, $1/\zeta_2 = 1$, and $n =$ 1,000 are used for calculations.

Estimates of the two constraints defined by Eq. (3.26) are listed in Table 3.5; parameters $d_1 = 6.9$, $d_2 = 0.27$, and $\bar{q} = 0.9$ were used for calculations. Note that estimates of $P(g_1(Y) \leq d_1)$ are identical under models m_1 and m_2, meaning that, assuming zero cost, if the utility function is based on values for $P(g_1(Y) \leq d_1|m_i)$, there will be no preference of one model over the other, *i.e.*, $m_1 \sim m_2$. However,

$m_1 \succ m_2$, *i.e.*, m_1 is superior to m_2, if we use values for $P(g_2(Y) \leq d_2|m_i)$ in the utility function since, as shown in Table 3.5, $P(g_2(Y) \leq d_2|m_1) > \bar{q}$ and $P(g_2(Y) \leq d_2|m_2) < \bar{q}$.

3.2.2 Continuous-time models

We next consider a class of continuous-time models for $Z(t)$. The design of the beam shown in Fig. 3.1 is then considered under the constraint given by Eq. (3.3), where the property of Y, denoted by $g(Y)$, is defined by Eq. (3.14) with $T = \{t : t \geq 0\}$.

3.2.2.1 Available information

No samples of the input are available; instead, we assume: (1) the second-moment properties of Z are given by

$$\begin{aligned} \mathrm{E}[Z(t)] &= 0, \\ \mathrm{E}[Z(t)Z(t+\tau)] &= \exp\left(-\theta \mid \tau \mid\right), \end{aligned} \tag{3.28}$$

where $\theta > 0$ is a deterministic constant, and (2) the marginal CDF of Z is lognormal, *i.e.*,

$$F(z) = \Phi\left[\frac{\ln\left(z-\alpha\right)}{\lambda}\right], \; z > \alpha, \tag{3.29}$$

where Φ denotes the standard normal CDF, and α, $\lambda > 0$ are deterministic constants. The corresponding marginal PDF for Z is

$$f(z) = \frac{\mathrm{d}F(z)}{\mathrm{d}z} = \frac{1}{\lambda(z-\alpha)}\,\phi\left[\frac{\ln\left(z-\alpha\right)}{\lambda}\right], \; z > \alpha, \tag{3.30}$$

where ϕ denotes the standard normal PDF. By Eqs. (3.28) and (3.29), the probability law for $Z(t)$ is partially known.

Table 3.5. Design performance under different model use.

	$P(g_1(Y) \leq d_1 \mid m_i)$	$P(g_2(Y) \leq d_2 \mid m_i)$
m_1	0.90	0.91
m_2	0.90	0.78

3.2.2.2 Candidate models

We consider the following class of candidate models for Z,

$$\mathcal{M}' = \{m_1, m_2\} = \{Z_1(t), Z_2(t)\}, \tag{3.31}$$

where model m_1 is a translation process and model m_2 is a diffusion process. It is shown in Sections 3.2.2.2.1 and 3.2.2.2.2 that if certain requirements are met, both processes are consistent with the available information on Z.

3.2.2.2.1 Translation model Let

$$Z_1(t) = F^{-1} \circ \Phi[G(t)] = \alpha + \exp\left[\delta G(t)\right], \tag{3.32}$$

where $G(t) \sim N(0, 1, c_G(\tau))$ is a zero-mean, unit variance, stationary Gaussian process with covariance function, $c_G(\tau)$. By Eq. (3.32), Z_1 is a monotonic, memoryless transformation of G, meaning that it is a **translation process** [47]. The parameters α and δ can be chosen so that $Z_1(t)$ has the target mean and variance from Eq. (3.28), *i.e.*, α and δ solve

$$\begin{aligned} \exp\left(\delta^2/2\right) + \alpha &= 0 \\ \exp\left(2\delta^2\right) - \alpha^2 &= 1. \end{aligned} \tag{3.33}$$

The covariance of $Z_1(t)$ is given by (see [48], p. 49)

$$c_{Z_1}(\tau) = \mathrm{E}[Z_1(t)Z_1(t+\tau)] = \exp\left[\delta^2(1 + c_G(\tau))\right] - \alpha^2. \tag{3.34}$$

For positive correlations, it can be shown that $c_{Z_1}(\tau) \approx c_G(\tau)$; this is further demonstrated by Fig. 3.10. Hence, we set $c_G(\tau) = \exp\left(-\theta|\tau|\right)$. The marginal distribution for $Z_1(t)$ is identical to the target CDF from Eq. (3.29) since

$$\begin{aligned} P(Z_1(t) \leq d) &= P\left(F^{-1} \circ \Phi[G(t)] \leq d\right) = P\left(G(t) \leq \Phi^{-1}[F(d)]\right) \\ &= F(d). \end{aligned} \tag{3.35}$$

3.2.2.2.2 Diffusion model Let Z_2 satisfy the following stochastic differential equation

$$\mathrm{d}Z_2(t) = -\theta Z_2(t)\mathrm{d}t + h\left[Z_2(t)\right]\mathrm{d}B(t), \tag{3.36}$$

where h is the diffusion term, and $B(t)$ is the Brownian motion. By Eq. (3.36), Z_2 is a **diffusion process** [88]. To enforce stationarity, the initial condition, $Z_2(t=0)$, is a random variable with CDF F. Using Itô's formula (see [48], Appendix H), it can be shown that the linear drift term in Eq. (3.36) guarantees an exponential correlation function, *i.e.*,

$$\mathrm{E}[Z_2(t)Z_2(s)] = \mathrm{E}\left[Z_2(t)^2\right]\exp\left[-\theta(s-t)\right],\ s \geq t. \tag{3.37}$$

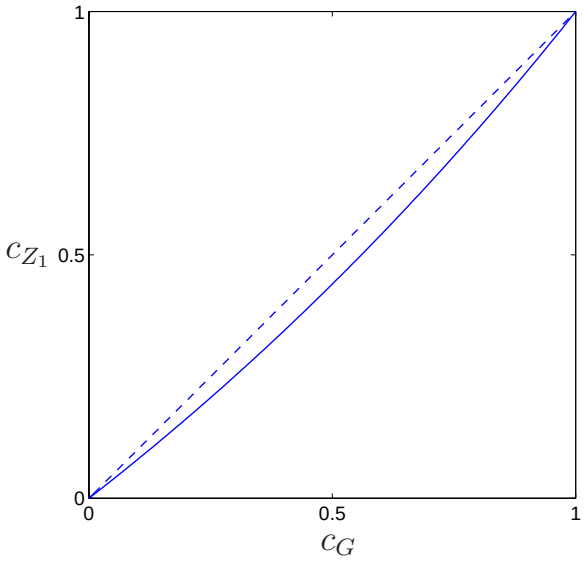

Figure 3.10: Covariance of translation process $Z_1(t)$ and its Gaussian image.

Assuming stationarity, the Fokker-Planck equation (see [50], p. 269) can be used to define h in terms of the target PDF, f,

$$h^2(z) = -\frac{2\theta}{f(z)} \int_{-\infty}^{z} v f(v) \mathrm{d}v. \tag{3.38}$$

Because $Z_2(t)$ follows the target marginal CDF, F, it can be shown that it also has the target mean and variance of Eq. (3.28).

3.2.2.3 Model selection using decision-theoretic method

We apply methods for model selection to: (1) choose a continuous-time model for Z, and (2) select a value for the design parameter, c, that satisfies the design requirement, *i.e.*, Eq. (3.3). Note that because data is unavailable, classical methods for model selection cannot be used (see Section 2.2). The decision-theoretic method for model selection is applied to choose the optimal design of the beam. Non-informative prior probabilities are assumed, so that models m_1 and m_2 are equally likely. The utility function is identical to the case of discrete-time models for Z considered in Section 3.2.1.3.2.

One sample of models m_1 and m_2 are shown in Fig. 3.11 for the case of $\theta = 1$. The samples look similar. Estimates of the PDF of the response of interest, $g(Y) = \max_{t \geq 0} |Y(t)|$, are shown in Fig. 3.12 under models m_1 and m_2, where it is evident that the translation model exhibits a slightly heavier tail. 1,000 Monte Carlo samples were generated for the analysis. The plot illustrates that even though the two models are equivalent in the second-moment sense, the distribution of $g(Y)$ is quite different.

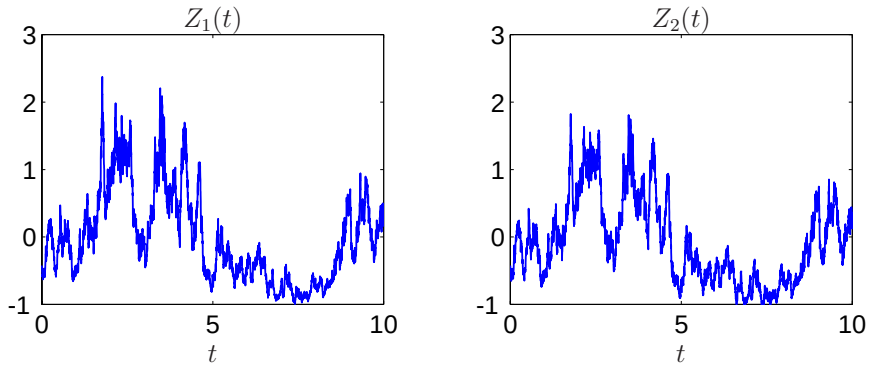

Figure 3.11: One sample of translation model, Z_1, and diffusion model, Z_2.

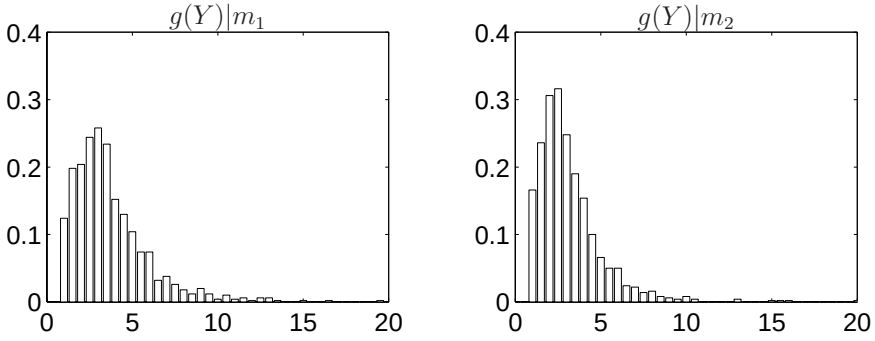

Figure 3.12: Histogram of $g(Y) = \max_t |Y(t)|$ under m_1 and m_2.

Table 3.6 shows the optimal model for three values of the model parameter, θ, and three values of the penalty function parameter, $\bar{\beta} = \beta_2/\beta_1$. The optimal model is sensitive to changes in both. For small values for $\bar{\beta}$, m_2 is optimal; as $\bar{\beta}$ increases, model m_1 is optimal. Note that for $\bar{\beta} = 1$, the penalty function, defined by Eq. (3.13), is symmetric so models for Z that result in conservative or non-conservative designs are penalized equally. As $\bar{\beta}$ increases, the asymmetry of ψ increases and models for Z that give a non-conservative design are assigned a large penalty. With a short correlation length (large θ), m_1 is optimal; as the correlation length increases, m_2 becomes optimal.

Table 3.6. Sensitivity of optimal model to changes in θ and $\bar{\beta}$.

θ	$\bar{\beta} = \frac{5}{2}$	$\bar{\beta} = 25$
0.3	m_2	m_2
1.0	m_2	m_1
2.5	m_1	m_1

3.3 Incomplete information on operator

Recall Eq. (3.1), where operator $\mathcal{D}$ with coefficient X maps input Z to output Y. We next consider the case where information on the functional form for $\mathcal{D}$ is incomplete. A new example is used for illustration since the functional form for $\mathcal{D}$ studied in Sections 3.1 and 3.2 is trivial.

Consider an oscillator with $q \geq 1$ degrees-of-freedom. The case of $q = 2$ is shown in Fig. 3.13, where $z(t)$ denotes the input, X denotes the value for a spring constant, and $w(t, X)$ denotes the relative displacement of the two masses. We assume: (i) the values for the two masses and one of the spring constants are known and fixed, (ii) X is a random variable uniformly distributed in $[\alpha, \beta]$, and (iii) input $z(t)$ is a perfectly known and deterministic function of time, t. Let random variable

$$Y = \max_{t \geq 0} \mid w(t, X) \mid \tag{3.39}$$

be the output of interest; the objective is to assess two metrics for performance: (1) $P(Y \leq d)$, where $d \geq 0$ is a deterministic parameter, and (2) E[Y]. We refer to these as performance metrics #1 and #2, respectively.

Recall Eq. (3.1), where $\mathcal{D}$: $(z(t), X) \rightarrow Y$. To estimate performance metrics #1 and #2, it is convenient to approximate $\mathcal{D}$; approximation may become necessary when we consider nonlinear systems or linear systems with many degrees-of-freedom,

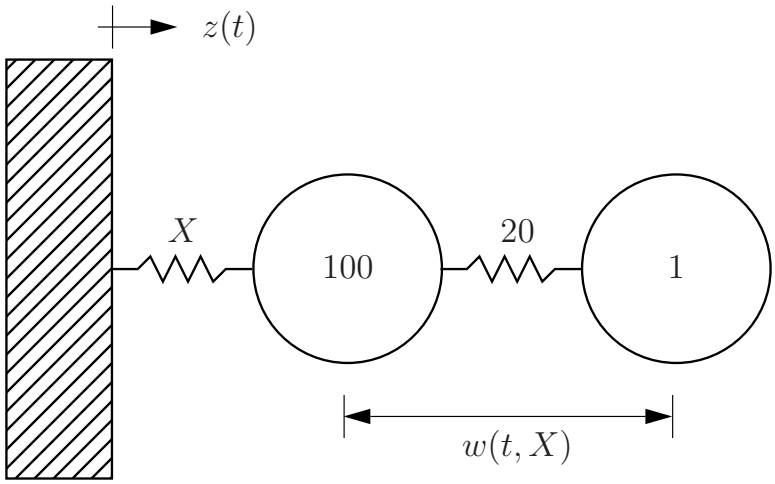

Figure 3.13: Two degree-of-freedom oscillator.

i.e., $q \gg 2$. A response surface (RS) model [64] for $\mathcal{D}$ is one viable approach. Herein, the decision-theoretic method for model selection is applied to: (i) select an optimal approximation for $\mathcal{D}$ from a class of candidate RS models, and (ii) assess the performance of the oscillator using metrics #1 and #2. By (ii), the model use is performance prediction.

3.3.1 Available information

Consider a fixed value for the stiffness, $x \in [\alpha, \beta]$, and let $\mathbf{m}$, $\mathbf{d}$, and $\mathbf{k}$ denote the 2×2 mass, damping, and stiffness matrices, respectively, of the two degree-of-freedom oscillator shown in Fig. 3.13, *i.e.*,

$$\mathbf{m} = \begin{pmatrix} 100 & 0 \\ 0 & 1 \end{pmatrix} \text{ and } \mathbf{k} = \begin{pmatrix} x+20 & -20 \\ -20 & 20 \end{pmatrix} \tag{3.40}$$

For calculations, we assume classical damping with a constant damping ratio of 4%. The relative displacement of the two masses, assuming zero initial conditions, is given by [88], p. 167,

$$w(t, x) = \mathbf{c} \int_0^t \exp\left[\mathbf{a}(t-\tau)\right] \mathbf{b} z(\tau) \mathrm{d}\tau, \tag{3.41}$$

where

$$\mathbf{a} = \begin{pmatrix} \mathbf{0} & \mathbf{i} \\ -\mathbf{m}^{-1}\mathbf{k} & -\mathbf{m}^{-1}\mathbf{d} \end{pmatrix} \mathbf{b} = \begin{pmatrix} 0 & 0 & -100 & -1 \end{pmatrix}^T \mathbf{c} = \begin{pmatrix} -1 & 1 & 0 & 0 \end{pmatrix} \tag{3.42}$$

and $\mathbf{i}$ denotes the 2×2 identity matrix. We solve Eq. (3.41) for $2 \leq n < \infty$ values for $x \in [\alpha, \beta]$, denoted by x_i, $i = 1, 2, \ldots, n$. In general, the accuracy of the RS

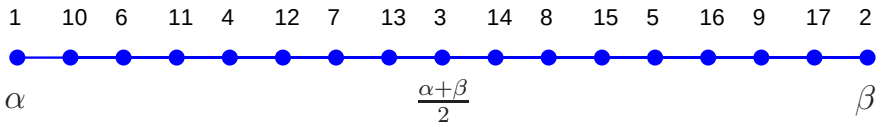

Figure 3.14: Values for x used to construct RS models for $\mathcal{D}$.

models depends on the values for x_i, $i = 1, 2, \ldots, n$. The points used here are shown in Fig. 3.14 for $n \leq 17$; the ordering of the points is also shown. For example, $x_3 = (\alpha + \beta)/2$. More general methods are discussed in [15].

For clarity, we define

$$v_i = \max_{t \geq 0} \mid w(t, x_i) \mid, \ i = 1, 2, \ldots, n. \tag{3.43}$$

The available information on $\mathcal{D}$ is given by (x_i, v_i), $i = 1, 2, \ldots, n$. The amount of available information increases with increasing n; the information on $\mathcal{D}$ is complete only when we solve Eq. (3.41) for each $x \in [\alpha, \beta]$.

3.3.2 Candidate models

Consider the following class of candidate models for $\mathcal{D}$

$$\begin{aligned} \mathcal{M}' &= \{m_1, m_2, m_3, m_4\} \\ &= \{y_1(x; \hat{\boldsymbol{\theta}}_1), y_2(x; \hat{\boldsymbol{\theta}}_2), y_3(x; \hat{\boldsymbol{\theta}}_3), y_4(x; \hat{\boldsymbol{\theta}}_4)\}, \end{aligned} \tag{3.44}$$

where

$$y_j(x; \hat{\boldsymbol{\theta}}_j) = \sum_{i=0}^{j} \hat{\theta}_{j,i} x^i, \ j = 1, 2, 3, 4. \tag{3.45}$$

Each member of $\mathcal{M}'$ is a RS model for $\mathcal{D}$; by Eq. (3.45), y_1, y_2, y_3, and y_4 are polynomials in x of degree one, two, three, and four, respectively. Note that $\mathcal{M}'$ is a function of n, *i.e.*, for $n = 2$ candidate models m_2, m_3, and m_4 are undefined, for $n = 3$, candidate models m_3 and m_4 are undefined, *etc.*; all candidate models are defined for $n \geq 5$. Under the available information, (x_i, v_i), $i = 1, 2, \ldots, n$, estimates of the model parameters, $\hat{\boldsymbol{\theta}}_j$, $j = 1, 2, 3, 4$, are calculated using the method of least squares [36]. We restrict the analysis to polynomial approximations for $\mathcal{D}$; more general functional forms are possible, *e.g.*, nonlinear regression and kriging [44].

The collection of candidate models is shown in Fig. 3.15 for the case of $n = 5$ (left) and $n = 9$ (right). The data, (x_i, v_i), $i = 1, 2, \ldots, n$, is also shown. Parameters $\alpha = 1{,}500$ and $\beta = 2{,}500$ were used for calculations; input $z(t)$ is one sample of zero mean Gaussian white noise. As shown, for $n = 5$ information is limited and the

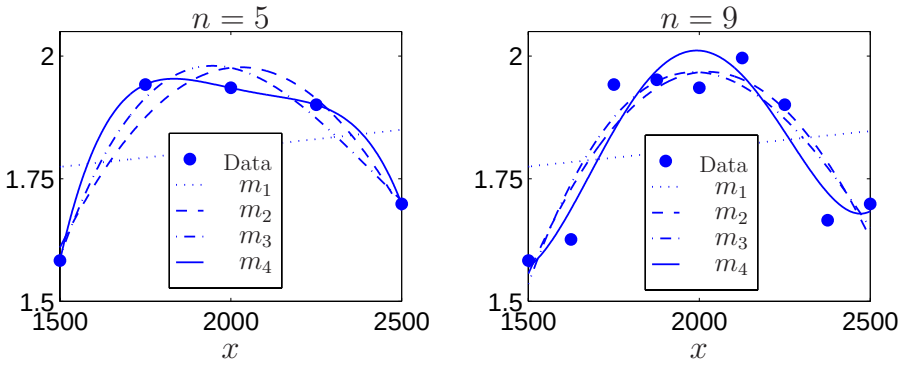

Figure 3.15: Candidate models for $n = 5$ and $n = 9$.

candidate models are very different; model m_4 passes thru all data points. For $n = 9$, m_2 and m_3 are similar, but no candidate model passes thru all data points.

We assume each member of $\mathcal{M}'$ can be expressed as a regression on the available information (see [1], Chapter 2):

$$V_{ij} = y_j(x_i; \hat{\boldsymbol{\theta}}_j) + E_{ij}, \ i = 1, 2, \ldots, n, \ j = 1, 2, 3, 4, \tag{3.46}$$

where, for each j, $\{E_{ij}\}$ is a sequence of zero-mean iid Gaussian random variables with variance σ_j^2. The likelihood of each candidate model is given by (see [98], p. 58)

$$l_j \propto \exp\left[-\frac{1}{2\hat{\sigma}_j^2}\sum_{k=1}^{n}\Big(v_k - y_j(x_k; \hat{\boldsymbol{\theta}}_j)\Big)^2\right], \ j = 1, 2, 3, 4, \tag{3.47}$$

where $\hat{\sigma}_j^2$ denotes an estimate for σ_j^2. The probability that model m_j is true, denoted by p_j, is the normalized likelihood of m_j given by Eq. (2.8). Recall that under the method of maximum likelihood, model m_i is optimal if $p_i \geq p_j$, $j = 1, 2, 3, 4$.

3.3.3 Model selection

The method of maximum likelihood and the decision-theoretic method for model selection developed in Chapter 2 are applied to: (i) select a model for $\mathcal{D}$, and (ii) assess the performance of the oscillator under metric #1, $P(Y \leq d)$, and metric #2, E[Y]. In general, different models for $\mathcal{D}$ will be selected for different methods of analysis and different metrics of performance.

The model use is performance prediction, so the formulation of Section 2.3.2.3 applies. Further, we consider the case where the collection of candidate actions, $\mathcal{A}$, coincides with the collection of candidate models, given by Eq. (3.44). The elements of $\mathcal{A}$ are therefore the candidate models, m_i, $i = 1, 2, 3, 4$. By Eq. (3.44), each $m_i \in \mathcal{M}'$ is deterministic, which illustrates that, as mentioned in Chapter 2, the decision-theoretic method for model selection does not require that the candidate models be random.

By Eq. (2.44), the utility is defined by the cost function, γ, and the penalty function, ψ. The cost of using each candidate model for prediction is identical, so without loss of generality we set $\gamma(m_i) = 0$, $i = 1, 2, 3, 4$. Under model m_j, a prediction of metric #1 is given by

$$h_{1,j} = P\left(Y \leq d \mid m_j\right) = \frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} 1\left(y_j(x; \hat{\boldsymbol{\theta}}_j) \leq d\right) \mathrm{d}x, \tag{3.48}$$

and a prediction of metric #2 is given by

$$\begin{aligned} h_{2,j} = \mathrm{E}\left[Y \mid m_j\right] &= \frac{1}{\beta - \alpha} \int_{\alpha}^{\beta} y_j(x; \hat{\boldsymbol{\theta}}_j) \mathrm{d}x \\ &= \frac{1}{\beta - \alpha} \sum_{i=0}^{j} \frac{\hat{\theta}_{j,i}}{i+1} \left(\beta^{i+1} - \alpha^{i+1}\right), \end{aligned} \tag{3.49}$$

where the last step follows since each $m_i \in \mathcal{M}'$ is a polynomial in x. For metric k, $k = 1, 2$, the penalty assigned to a prediction made under model m_i, if model m_j is true, is given by

$$\psi(m_i, m_j) = \tilde{\psi}(h_{k,i}, h_{k,j}) = \begin{cases} \beta_1 (h_{k,i} - h_{k,j})^2 & \text{if } h_{k,i} \leq h_{k,j}, \text{ and} \\ \beta_2 (h_{k,i} - h_{k,j})^2 & \text{if } h_{k,i} > h_{k,j}, \end{cases} \tag{3.50}$$

where $\beta_2 > \beta_1 \geq 0$ are deterministic parameters. Equation (3.50) is identical to Eq. (2.73) defined in Section 2.3.2.3.

The expected utility of each candidate model is given by Eq. (2.45); the optimal model, denoted by $m^\star$, minimizes Eq. (2.45). Optimal models under various values for n are shown in Fig. 3.16. Shown at the top of Fig. 3.16 is $m^\star$ under the method of maximum likelihood. Results using the decision-theoretic method for model selection are shown at the bottom of Fig. 3.16, illustrating that the optimal model depends on the model use; $m^\star$ for performance metric #1 is shown on the left, while $m^\star$ for performance metric #1 is shown on the right. The method of maximum likelihood is independent of model use. In each case, for $1 < n \leq 5$, we have $m^\star = m_{n-1}$; for $n > 5$, different models are selected in the three plots. Under the method of maximum likelihood, $m^\star = m_4$ for $n \geq 5$, indicating that the highest-order RS approximation is always optimal. However, predictions of performance metrics #1 and #2 using model m_4 are non-conservative for certain values for $n \geq 5$; the decision-theoretic method for model selection accounts for this.

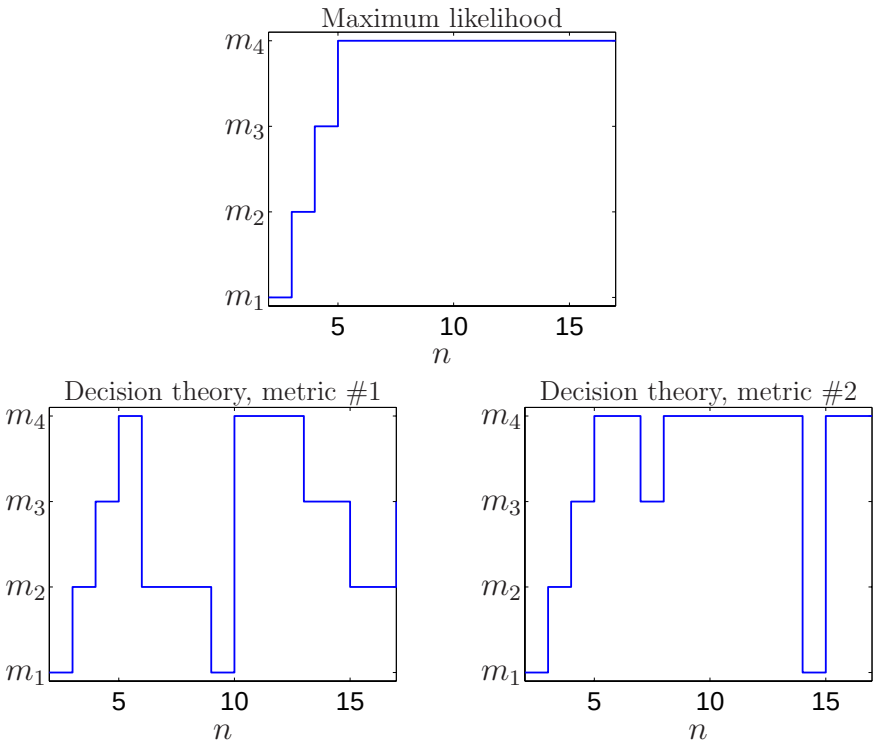

Figure 3.16: Optimal model, $m^\star$, *vs.* amount of information, n.

This page intentionally left blank.

Chapter 4

The polynomial chaos approximation

Polynomial chaos (PC) representations [43, 60, 94] for non-Gaussian random variables and stochastic processes are infinite series of Hermite polynomials of standard Gaussian random variables with deterministic coefficients. These series converge in mean square to the functions they replace, provided these functions are in L_2. PC representations are a topic of continued research because, among other features, they provide a framework suitable for computational simulation.

For calculations, the PC representations are truncated, creating what is herein referred to as PC approximations. The use of PC approximations is widespread. Applications include structural fatigue [95], structural reliability [100], structural mechanics [41, 43], linear structural dynamics [82], nonlinear random vibration [42], solution of stochastic differential equations [6], soil mechanics [40], soil-structure interaction [72], and simulation of non-Gaussian random fields [39, 77, 78, 86]. Generally, the PC approximations used in applications have ten or fewer terms [39, 41, 42, 86].

Features and limitations of the PC approximation are discussed in Section 4.1, where it is shown that the accuracy of the PC approximation depends on the number of terms retained in the truncated series. The decision-theoretic method for model selection is used to determine the optimal number of terms to retain in Section 4.2.

4.1 Features and limitations of polynomial chaos

Properties of PC approximations are explored by using simple but relevant examples. It is shown that: (1) the rate of convergence of PC approximations may be slow, (2) PC approximations may not improve by adding terms, (3) moments of order three and higher calculated for PC approximations may be inaccurate, and (4) PC

approximations for stationary, non-Gaussian processes may not be stationary. These observations raise questions about the use of PC approximations, particularly for problems requiring accurate tail probabilities, *e.g.*, reliability analysis. With all these limitations, PC approximations can still provide useful results in many applications, as demonstrated by some of the above references.

4.1.1 Random vectors

Consider the class of non-Gaussian, $\mathbb{R}^d$-valued random variables, $\mathbf{Y}$, that can be written as a function of a standard Gaussian random vector, *i.e.*,

$$\mathbf{Y} = \mathbf{g}(\mathbf{W}), \tag{4.1}$$

where $\mathbf{g}\colon \mathbb{R}^k \to \mathbb{R}^d$ is a deterministic, measurable mapping, $\mathbf{W} \sim N(\mathbf{0}, \mathbf{i})$ is an $\mathbb{R}^k$-valued vector of independent, identically distributed (iid), zero mean, unit variance Gaussian random variables, and $\mathbf{i}$ denotes the (k, k) identity matrix. It is assumed that $\mathbf{Y}$ is in L_2, so that all coordinates of $\mathbf{Y}$ have finite second moment. For example, in the case of scalar random variables Y and W, $Y = F_Y^{-1} \circ \Phi(W) = g(W)$ follows the distribution F_Y, where $\Phi(\,\cdot\,)$ denotes the standard normal cumulative distribution function (CDF) and $\mathrm{E}[g(W)^2]$ exists and is finite.

4.1.1.1 PC representation for random vectors

Under the assumption that the j^{th} coordinate of $\mathbf{Y}$ has finite variance, the series [43, 60]

$$Y_j = g_j(\mathbf{W}) = \sum_{i_1, i_2, \ldots, i_k \geq 0} y^{(j)}_{i_1, i_2, \ldots, i_k} h_{i_1, i_2, \ldots, i_k}(\mathbf{W}),\ j = 1, 2, \ldots, d, \tag{4.2}$$

is convergent in L_2 and constitutes the polynomial chaos (PC) representation for Y_j. Here, $y^{(j)}_{i_1, i_2, \ldots, i_k}$, $j = 1, 2, \ldots, d$, are deterministic coefficients that must be determined, and $h_{i_1, i_2, \ldots, i_k}$ are k-dimensional orthogonal Hermite polynomials given by [48], p. 420

$$h_{\alpha, \ldots, \omega}(\mathbf{W}) = e^{\frac{1}{2}\mathbf{W}^T\mathbf{W}} \left(-\frac{\partial}{\partial W_\alpha}\right) \cdots \left(-\frac{\partial}{\partial W_\omega}\right) e^{-\frac{1}{2}\mathbf{W}^T\mathbf{W}}, \tag{4.3}$$

where $\mathbf{W} = [W_1 W_2 \ldots W_k]^T$ and $\alpha, \ldots, \omega$ are indices in $\{0, \ldots, k\}$ that need not be different. The Hermite polynomials have the properties

$$\begin{aligned} \mathrm{E}\left[h_{\alpha_i, \ldots, \omega_i}(\mathbf{W}) h_{\alpha_j, \ldots, \omega_j}(\mathbf{W})\right] &= \mathrm{E}\left[h_{\alpha_i, \ldots, \omega_i}(\mathbf{W})^2\right] \delta_{ij}, \\ \mathrm{E}\left[h_{\alpha_i, \ldots, \omega_i}(\mathbf{W})\right] &= \begin{cases} 1, & i = 0, \\ 0, & \forall i \geq 1, \end{cases} \end{aligned} \tag{4.4}$$

and define an orthogonal basis in L_2 [41]. The coefficients of the series in Eq. (4.2) can be calculated by exploiting the properties of Eq. (4.4). For example, the PC representation in Eq. (4.2) simplifies to

$$Y = g(W_1, W_2) = \sum_{i_1,i_2=0}^{\infty} y_{i_1,i_2} h_{i_1,i_2}(W_1, W_2) \tag{4.5}$$

for $d = 1, k = 2$, where $y^{(1)}_{i_1,i_2} = y_{i_1,i_2}$, and the first few Hermite polynomials of Eq. (4.3) are

$$\begin{aligned} h_{0,0}(W_1, W_2) &= 1, \\ h_{1,0}(W_1, W_2) &= W_1, \\ h_{0,1}(W_1, W_2) &= W_2, \\ h_{2,0}(W_1, W_2) &= W_1^2 - 1, \\ h_{1,1}(W_1, W_2) &= W_1 W_2, \text{ and} \\ h_{0,2}(W_1, W_2) &= W_2^2 - 1. \end{aligned} \tag{4.6}$$

The coefficients of the series in Eq. (4.5) result from

$$\begin{aligned} \mathrm{E}\left[Y h_{j_1,j_2}(W_1, W_2)\right] &= \mathrm{E}\left[\sum_{i_1,i_2=0}^{\infty} y_{i_1,i_2} h_{i_1,i_2}(W_1, W_2) h_{j_1,j_2}(W_1, W_2)\right] \\ &= \sum_{i_1,i_2=0}^{\infty} y_{i_1,i_2} \mathrm{E}\left[h_{i_1,i_2}(W_1, W_2) h_{j_1,j_2}(W_1, W_2)\right] \\ &= y_{j_1,j_2} \mathrm{E}\left[h_{j_1,j_2}(W_1, W_2)^2\right], \end{aligned} \tag{4.7}$$

provided that the expectation can be calculated term by term. Under this condition, the coefficients in Eq. (4.2) are given by

$$y^{(j)}_{i_1,i_2,\ldots,i_k} = \frac{\mathrm{E}\left[Y_j h_{j_1,j_2,\ldots,j_k}(\mathbf{W})\right]}{\mathrm{E}\left[h_{j_1,j_2,\ldots,j_k}(\mathbf{W})^2\right]},\ j = 1, 2, \ldots, d. \tag{4.8}$$

4.1.1.2 PC approximation for random vectors

Let

$$Y_j^{(p)} = g_j^{(p)}(\mathbf{W}) = \sum_{\substack{i_1,i_2,\ldots,i_k \geq 0 \\ i_1+i_2+\cdots+i_k \leq q}} \tilde{y}^{(j)}_{i_1,i_2,\ldots,i_k} h_{i_1,i_2,\ldots,i_k}(\mathbf{W}),\ j = 1, 2, \ldots, d, \tag{4.9}$$

be the PC approximation for Y_j in Eq. (4.2), which consists of Hermite polynomials up to and including order q. The PC approximation has $p + 1$ terms, where [43]

$$p = \sum_{s=1}^{q} \frac{1}{s!} \left\{ \prod_{r=0}^{s-1} (k + r) \right\}. \tag{4.10}$$

The coefficients $\tilde{y}^{(j)}_{i_1,i_2,\ldots,i_k}$ of the PC approximation in Eq. (4.9) are taken to be equal to the coefficients $y^{(j)}_{i_1,i_2,\ldots,i_k}$ of the PC representation in Eq. (4.2). For example, the PC approximation for the series in Eq. (4.5) is

$$Y^{(p)} = \sum_{\substack{i_1,i_2 \geq 0 \\ i_1+i_2 \leq q}} y_{i_1,i_2} h_{i_1,i_2}(W_1, W_2). \tag{4.11}$$

The PC approximation has two useful properties: (1) $\mathrm{E}[Y^{(p)}] = \mathrm{E}[Y]$ for any p, and (2) $Y_j^{(p)} \xrightarrow[p\to\infty]{L_2} Y_j$. The latter is useful since it guarantees that the PC approximation of Eq. (4.9) approaches the PC representation of Eq. (4.2) in mean square as the number of terms retained, p, increases. However, this property is of limited use for applications since, generally, p is small [39, 41, 42, 86] so that asymptotic properties of $Y_j^{(p)}$ do not apply.

An additional approximation in applications arises because, in general, $\mathbf{g}(\mathbf{W})$ is not known explicitly. In such cases, the right-hand side of Eq. (4.8) must be estimated, causing a further layer of approximation [32]. For example, consider $\mathbf{Y}$ to be the output from a finite element model. In this case, only realizations of $\mathbf{g}(\mathbf{W})$ are available in the form of runs of the finite element code; the functional form of $\mathbf{g}(\mathbf{W})$ is not known and, therefore, the numerator of the right-hand side of Eq. (4.8) cannot be solved in closed-form. These issues are not addressed here.

The accuracy of the PC approximation in Eq. (4.9) is assessed by three examples in which $d = k = 1$, the map $Y = g(W)$ is known explicitly, and the right-hand side of Eq. (4.8) can be solved in closed-form. Hence, only the error associated with the truncation of the infinite series at p terms remains.

4.1.1.3 Accuracy of PC approximation

Seven metrics are used to evaluate the accuracy of the PC approximation. These metrics are listed in Table 4.1. The first metric is the difference between the exact and approximate mappings of W used to define Y and $Y^{(p)}$; this difference is calculated at $W = 0$ and $W = 4$. Metric two is the scaled mean-square error of the PC approximation. The probability that the absolute value of the difference between Y and $Y^{(p)}$ exceeds a specified $\epsilon > 0$ is the third metric, where Ω denotes the sample space of Y and $Y^{(p)}$. The metric can be interpreted as a measure of the order, p, required for $Y^{(p)}$ to equal Y in probability. Metric four can be viewed as the probability that $Y^{(p)}$ takes values that are not physically acceptable, while $\mu_5(p)$ provides a measure for the difference between the tails of Y and $Y^{(p)}$ at the 1% upper fractile. Metrics $\mu_6(p)$ and $\mu_7(p)$ compare the variance and kurtosis, denoted by $\mathrm{Var}[\,\cdot\,]$ and $\mathrm{Kur}[\,\cdot\,]$ respectively, of $Y^{(p)}$ to Y. Each of these metrics is computed for $1 \leq p \leq 20$, the upper limit on p being more terms than is typically used for applications.

Table 4.1. Metrics used to assess the accuracy of the PC approximation.

	Metric
$\mu_1(p)$	$\begin{cases} \frac{Y-Y^{(p)}}{Y}, & \text{if } Y \neq 0; \\ Y - Y^{(p)}, & \text{otherwise,} \end{cases}$
$\mu_2(p)$	$\dfrac{\mathrm{E}\left[(Y-Y^{(p)})^2\right]}{\mathrm{E}\left[Y^2\right]}$
$\mu_3(p)$	$P\left(G\right), G = \{\omega \in \Omega \colon \lvert Y(\omega) - Y^{(p)}(\omega)\rvert > \epsilon\}, \epsilon > 0$
$\mu_4(p)$	$P\left(\left\{Y^{(p)} \leq \inf_{\omega\in\Omega} Y\right\} \cup \left\{Y^{(p)} \geq \sup_{\omega\in\Omega} Y\right\}\right)$
$\mu_5(p)$	$P\left(Y^{(p)} > F_Y^{-1}(0.99)\right)$
$\mu_6(p)$	$\dfrac{\mathrm{Var}\left[Y^{(p)}(W)\right]}{\mathrm{Var}\left[Y(W)\right]}$
$\mu_7(p)$	$\dfrac{\mathrm{Kur}\left[Y^{(p)}(W)\right]}{\mathrm{Kur}\left[Y(W)\right]}$

Table 4.2. Random variable examples: (1) lognormal, (2) uniform on $[\alpha, \beta]$, and (3) reflected Gaussian.

Example	Mapping, $g(\cdot)$ / CDF, $F_Y(\cdot)$	Coefficients, y_i
(1)	$Y_1 = \alpha + \exp(\beta W)$	$y_0 = \alpha + e^{\beta^2/2}$ $y_j = \dfrac{\beta^j}{j!} e^{\beta^2/2},\ j = 1, 2, \ldots$
(2)	$Y_2 = F_{Y_2}^{-1} \circ \Phi(W)$, where $F_{Y_2}(y) = \begin{cases} 0, & y < \alpha; \\ \frac{y-\alpha}{\beta-\alpha}, & y \in [\alpha, \beta); \\ 1, & y \geq \beta. \end{cases}$	$y_0 = (\alpha + \beta)/2$ $y_{2j} = 0$ $y_{2j+1} = \dfrac{(-1)^{j+1}(\alpha - \beta)}{2^{j+1}\sqrt{\pi}(2j+1)!}\eta(j),\ j = 0, 1, \ldots$ $\eta(j) = \lvert -1 \cdot 1 \cdot 3 \cdots (2j-1) \rvert$
(3)	$Y_3 = \lvert W \rvert$	$y_{2j} = \dfrac{2}{\sqrt{2\pi}} \left\{ \dfrac{2^j j!}{(2j)!} + \displaystyle\sum_{i=1}^{j} (-1)^i \dfrac{2^{j-2i}(j-i)!}{i!(2j-2i)!} \right\}$ $y_{2j+1} = 0,\ j = 0, 1, \ldots$

4.1.1.4 Examples

Let $Y_1 = \alpha + \exp(\beta W)$, $Y_2 = F_{Y_2}^{-1} \circ \Phi(W)$, and $Y_3 = |W|$ be three random variables, where $W \sim N(0,1)$, $\alpha, \beta > 0$ are deterministic constants, and F_{Y_2} is the CDF of a random variable uniformly distributed in $[\alpha, \beta]$. Table 4.2 lists the definitions of Y_1, Y_2, and Y_3, as well as the PC coefficients given by Eq. (4.8). The derivation of the PC coefficients for Y_3 is included as Appendix A.

The PC coefficients for random variables Y_1, with $(\alpha = 0, \beta = 1)$, Y_2, with $(\alpha = -1, \beta = 1)$, and Y_3 are plotted in Fig. 4.1, showing they all approach zero for increasing p. These coefficients decrease monotonically with p only for Y_1. The coefficients y_{2k} for Y_2 and y_{2k+1} for Y_3, $k = 0, 1, 2, \ldots$, are zero.

Shown in Figs. 4.2-4.4 are the true maps, $g_i(W)$, and approximate maps, $g_i^{(p)}(W)$, for Y_i, $i = 1, 2, 3$, and $p = 1, 2, \ldots, 9$. In general, as more terms are retained in Eq. (4.9), each approximation improves. For $Y_1^{(p)}$, the map is accurate near $W = 0$ with $p = 3$, but $p > 5$ terms are needed for $g_1^{(p)}$ to approximate g_1 satisfactorily for all $W \in [-4, 4]$. For $p = 9$, it is difficult to distinguish $g_1^{(p)}(W)$ from $g_1(W)$. The mapping $g_2^{(p)}$ approximates g_2 satisfactorily at $W = 0$ with few terms, but is less satisfactory in the vicinity of $W = \pm 4$, even for large p. The mapping $g_3^{(p)}$ requires more terms than $g_1^{(p)}(W)$ or $g_2^{(p)}(W)$ to be satisfactory in $[-4, 4]$. It has notable oscillations in a vicinity of $W = \pm 4$ and approaches zero slowly at $W = 0$.

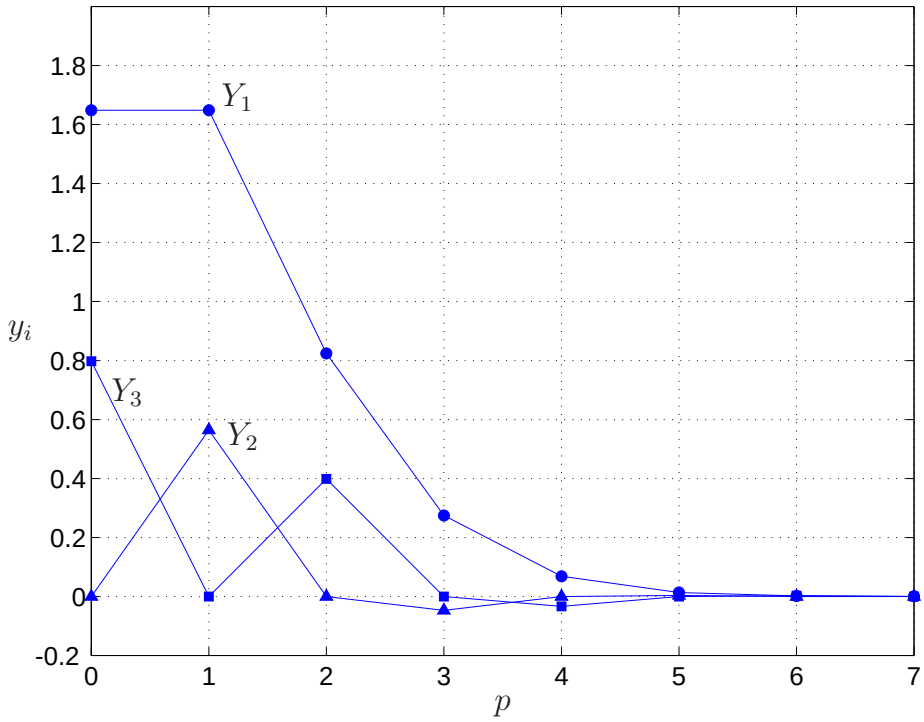

Figure 4.1: PC coefficients for Y_1, Y_2, and Y_3.

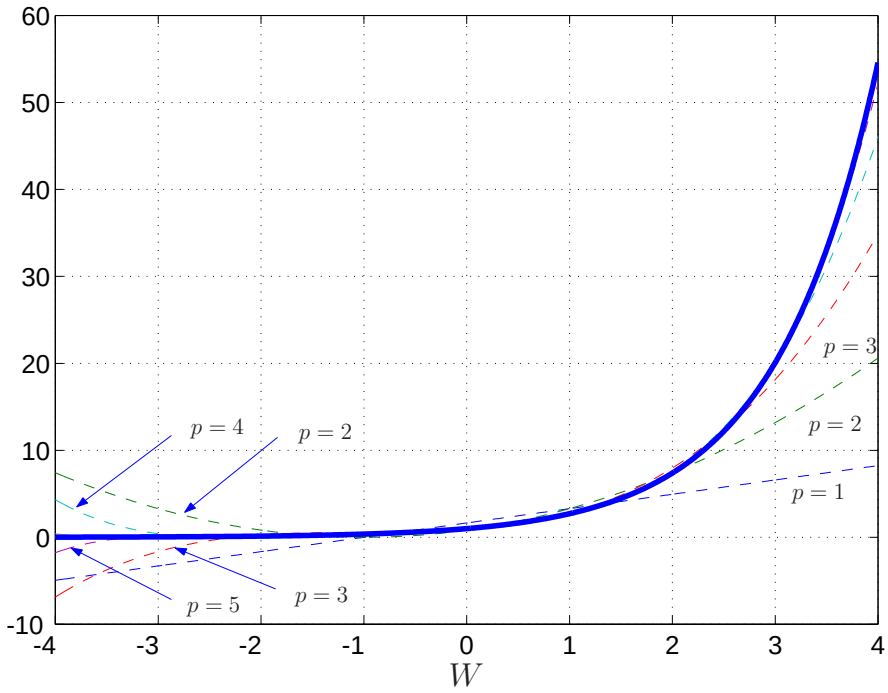

Figure 4.2: Exact map, $Y_1 = g_1(W)$, *vs.* approximate map, $Y_1^{(p)} = g_1^{(p)}(W)$.

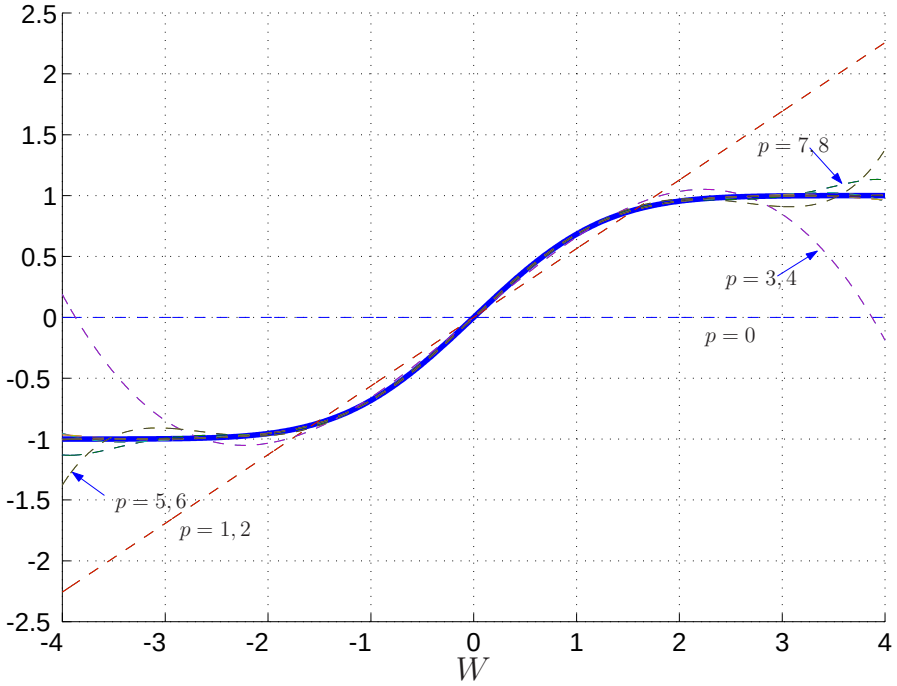

Figure 4.3: Exact map, $Y_2 = g_2(W)$, *vs.* approximate map, $Y_2^{(p)} = g_2^{(p)}(W)$.

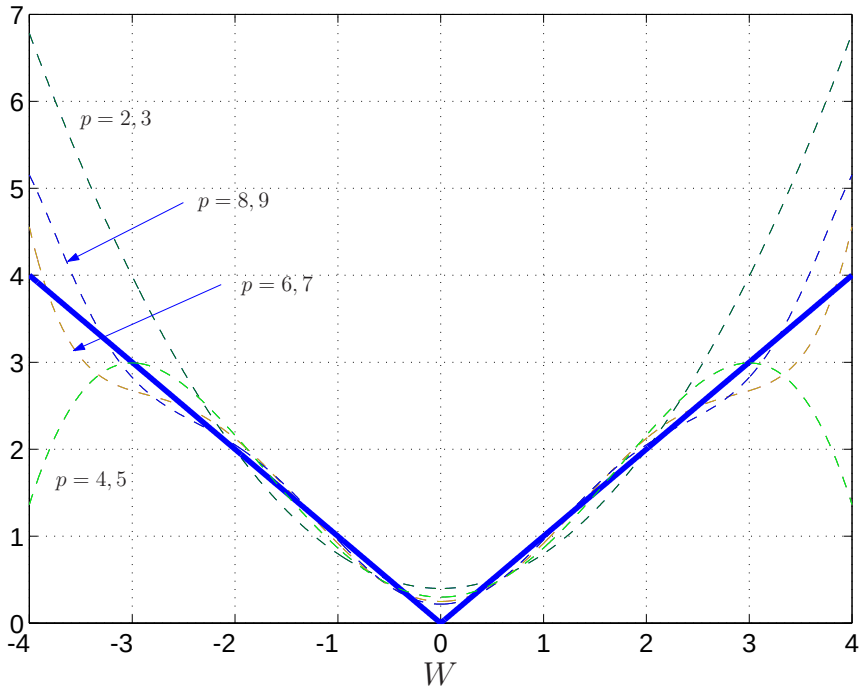

Figure 4.4: Exact map, $Y_3 = g_3(W)$, *vs.* approximate map, $Y_3^{(p)} = g_3^{(p)}(W)$.

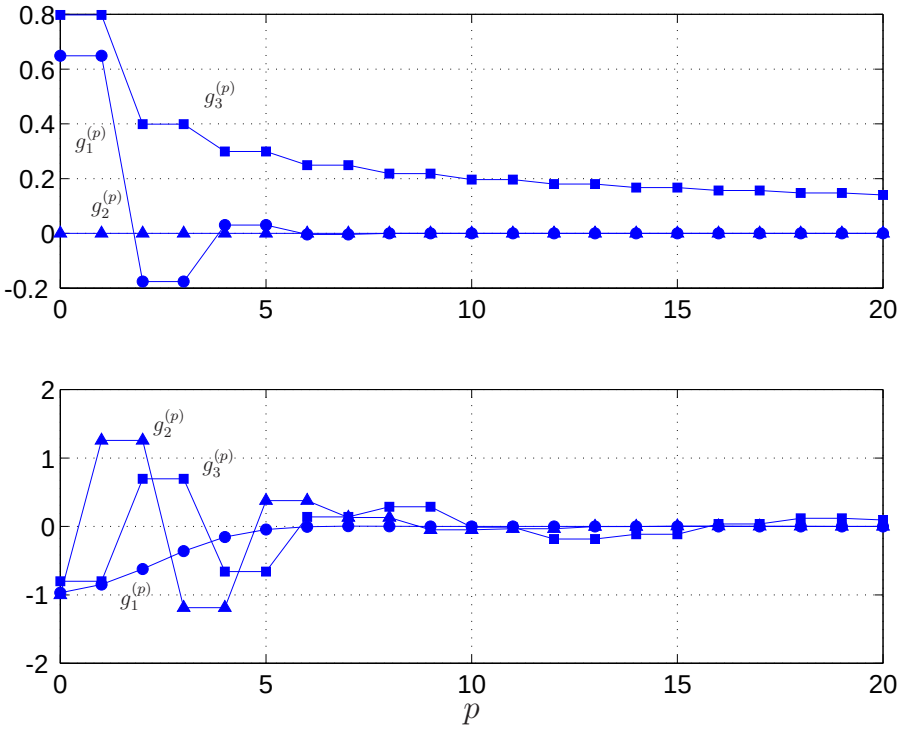

Figure 4.5: Metric #1: PC maps at $W = 0$ (top) and $W = 4$ (bottom).

The accuracy of the approximate maps is further illustrated in Fig. 4.5, showing plots of $\mu_1(p)$ for each example. For $g_1^{(p)}$, $\mu_1(p)$ is small for both $W = 0$ and $W = 4$ when $p \geq 6$. $g_2^{(p)}$ is accurate at $W = 0$ for all p, but exhibits oscillatory behavior at $W = 4$. The oscillations are negligible for $p > 13$. At $W = 0$, $g_3^{(p)}$ with $p = 20$ exhibits a 20% error, while at $W = 4$, $g_3^{(p)}$ exhibits significant oscillatory behavior at $p = 20$. Using the PC approximation for the lognormal random variable, $g_1^{(p)}(W)$, as a baseline for comparison, the approximation for the uniform random variable, $g_2^{(p)}(W)$, exhibited slower convergence and noise in the tails. The approximation for the reflected Gaussian $g_3^{(p)}(W)$ is unsatisfactory.

Figure 4.6 shows $\mu_2(p)$, the error defined in the second row of Table 4.1, as a function of p. Results have been obtained by Monte Carlo simulation using 100,000 samples. The mean square error decays most rapidly for $Y_1^{(p)}$, followed by the mean square errors for $Y_2^{(p)}$ and $Y_3^{(p)}$. Estimates of $\mu_3(p)$ are plotted in Fig. 4.7, where a similar pattern is observed. Y_1 and $Y_1^{(p)}$ can be viewed as equal in probability for $p \geq 10$, while Y_2 and $Y_2^{(p)}$ for $p \geq 18$. The tails of $Y_3^{(p)}$ and Y_3 differ significantly, even for $p = 20$. The calculations were performed with $\epsilon_i = 0.001 \cdot \sqrt{\text{Var}[Y_i]}$, $i = 1, 2, 3$.

The random variables Y_1 with $(\alpha = 0, \beta = 1)$, Y_2 with $(\alpha = -1, \beta = 1)$, and Y_3 are such that

$$Y_1, Y_3 \geq 0, \; Y_2 \in [-1, 1] \tag{4.12}$$

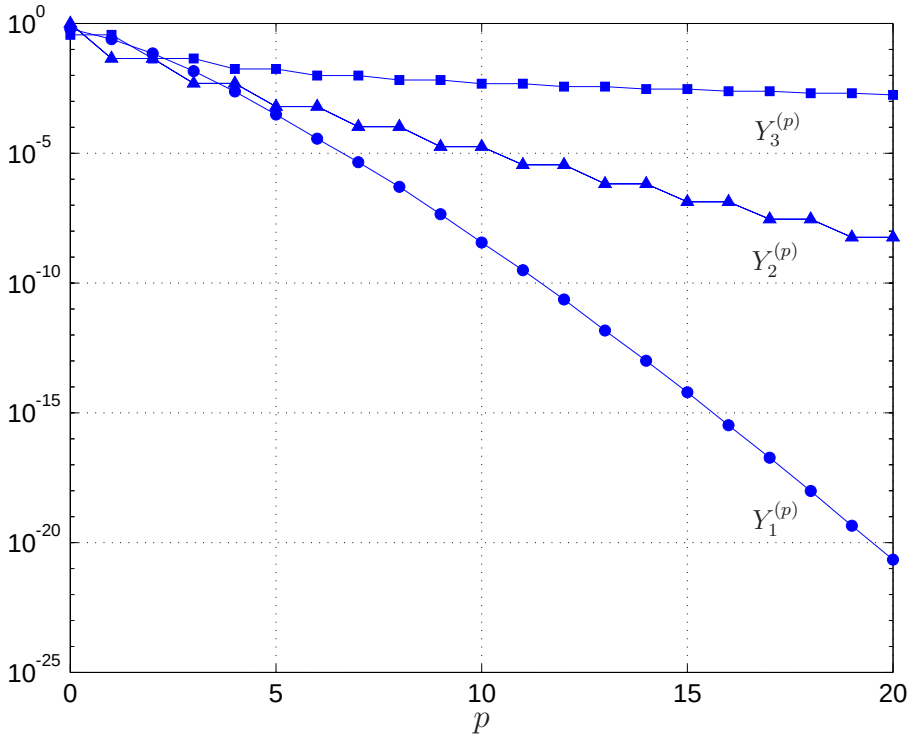

Figure 4.6: Metric #2: estimates of relative mean-square error.

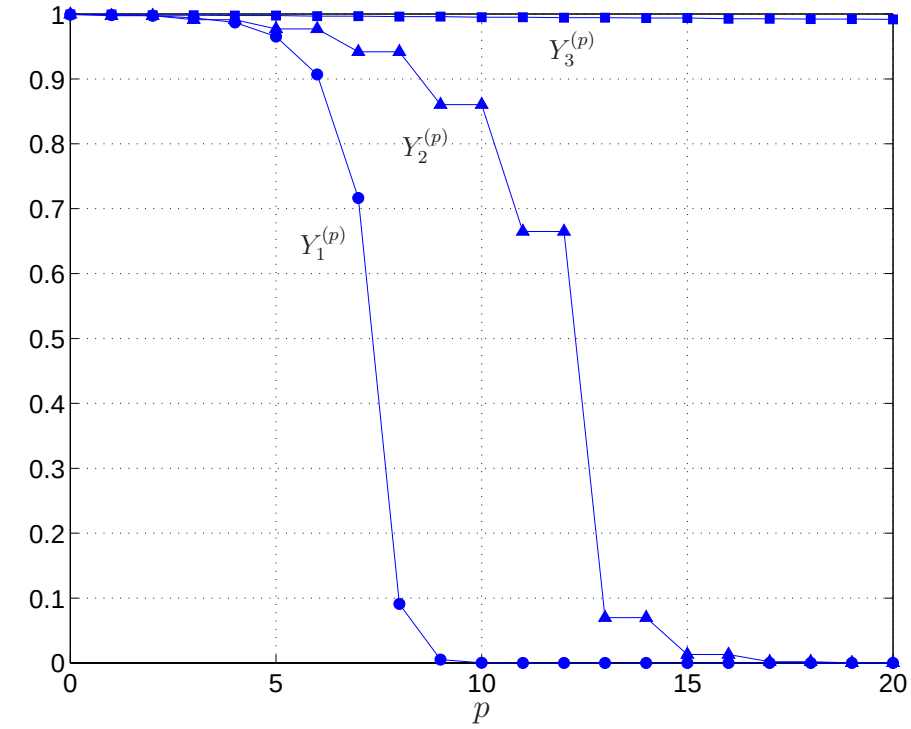

Figure 4.7: Metric #3: estimates of $P(G)$.

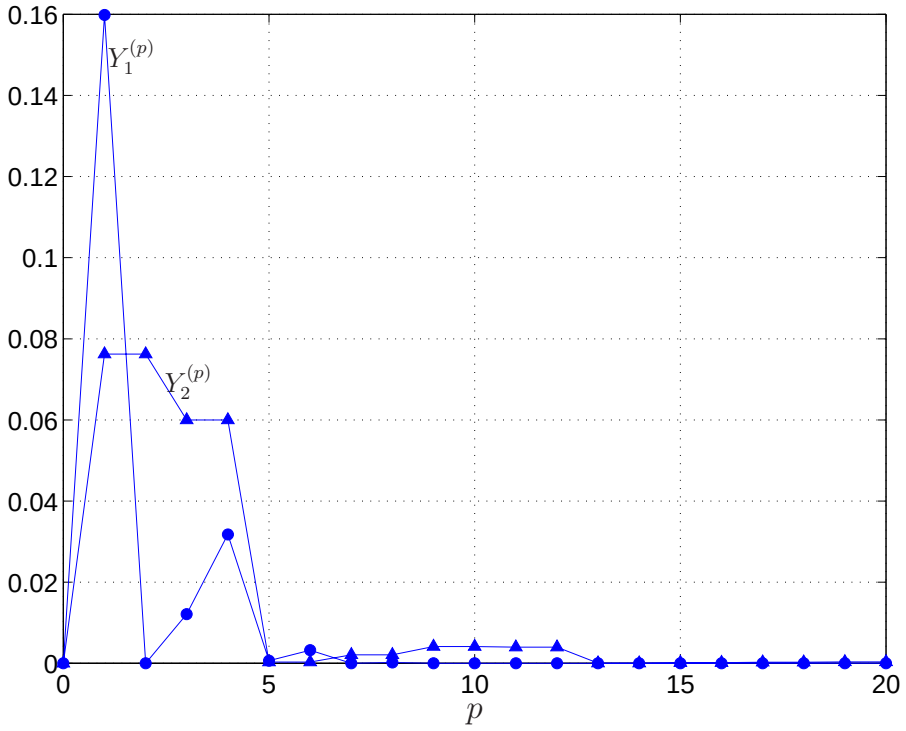

Figure 4.8: Metric #4: estimates of $P(Y_1^{(p)} \leq 0)$ and $P(\{Y_2^{(p)} \leq -1\} \cup \{Y_2^{(p)} \geq 1\})$.

almost surely. Estimates of $P(Y_1^{(p)} \leq 0)$, $P(\{Y_2^{(p)} \leq -1\} \cup \{Y_2^{(p)} \geq 1\})$, and $P(Y_3^{(p)} \leq 0)$ are shown in Fig. 4.8; results have been obtained using 100,000 samples. While $P(Y_1^{(p)} \leq 0)$ is initially high, it rapidly decays to near zero. $P(Y_2^{(p)} \leq -1) \cup \{Y_2^{(p)} \geq 1\})$ is initially lower, but decays less rapidly; it remains at 0.001 for $p = 20$. $P(Y_3^{(p)} \leq 0) = 0$ for all p because, as shown by the coefficients of Table 4.2, all odd coefficients (odd powers of W) in the PC approximation are zero.

Shown in Fig. 4.9 is $\mu_5(p)$, the 1%-upper fractile estimates for $Y_1^{(p)}$, $Y_2^{(p)}$, and $Y_3^{(p)}$; results have been obtained using 100,000 samples. The PC approximation for the lognormal random variable is satisfactory for small p, while the PC approximation to the uniform random variable requires all terms considered before satisfactory performance is achieved. The PC approximation to the reflected Gaussian random variable exhibits a large error at $p = 20$.

The two remaining deterministic metrics, the normalized variance and kurtosis of the PC approximations, are shown in Figs. 4.10 and 4.11, respectively. The normalized variance of $Y_1^{(p)}$ and $Y_2^{(p)}$ are nearly one for $p \geq 7$. The normalized variance of $Y_3^{(p)}$ is in error by less than 1% at $p = 20$. The kurtosis coefficient of $Y_1^{(p)}$ and Y_1 practically coincide for $p \geq 10$, but the kurtosis of $Y_3^{(p)}$ diverges from the kurtosis of Y_3 as p increases. It is not clear whether $\mu_7(p)$ for $Y_2^{(p)}$ converges to one or not as p increases. Note that $\mu_6(p)$ and $\mu_7(p)$ can be calculated directly; no sampling is required.

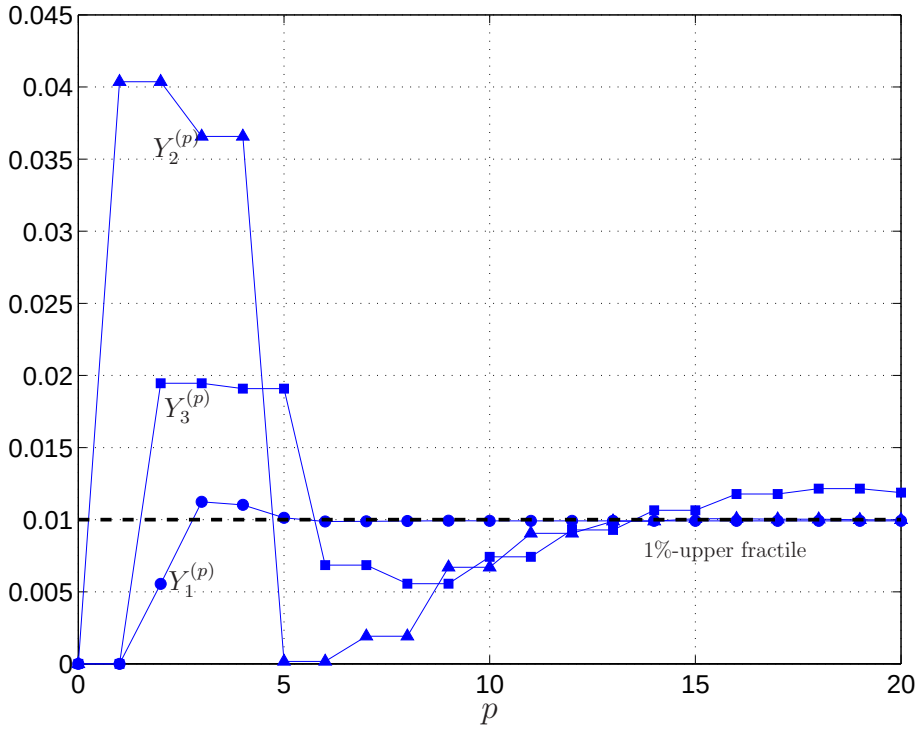

Figure 4.9: Metric #5: Estimates of 1%-upper fractile.

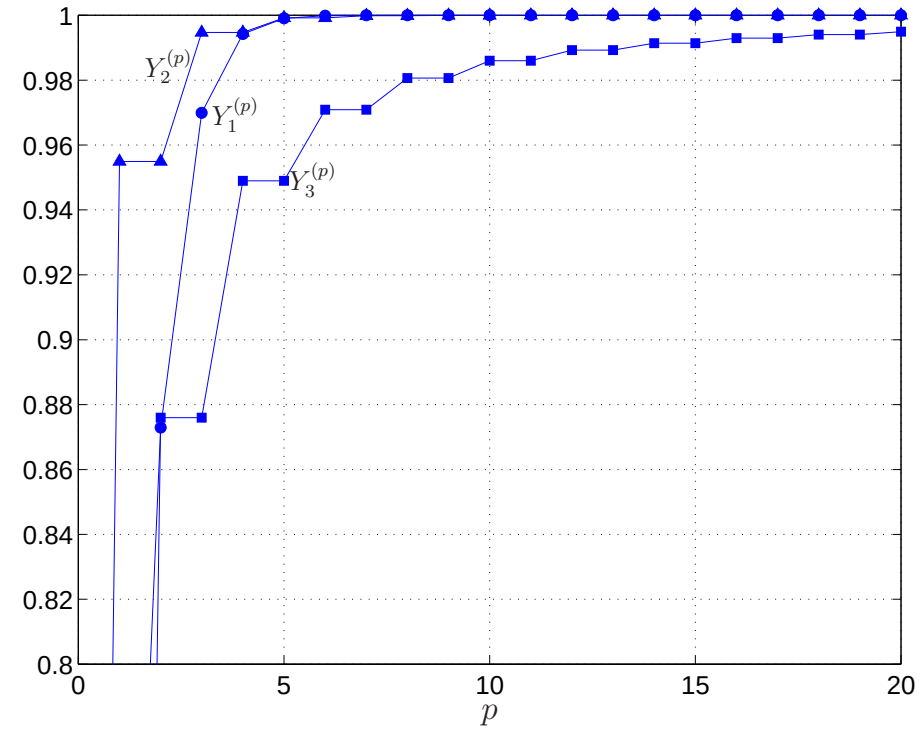

Figure 4.10: Metric #6: normalized variance of $Y_1^{(p)}$, $Y_2^{(p)}$, and $Y_3^{(p)}$.

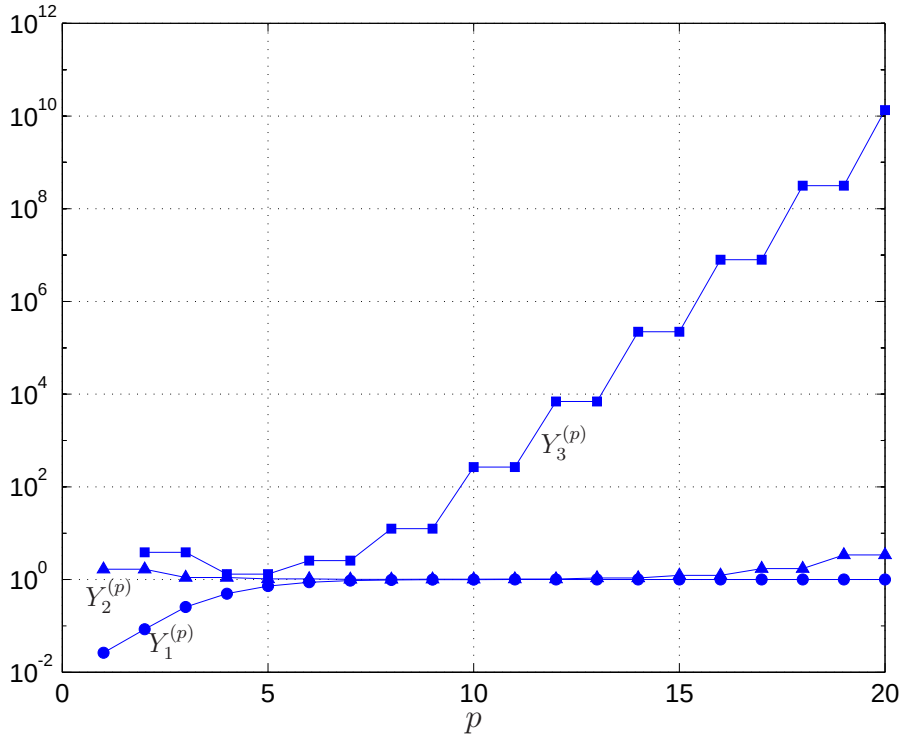

Figure 4.11: Metric #7: normalized kurtosis of $Y_1^{(p)}$, $Y_2^{(p)}$, and $Y_3^{(p)}$.

The rate of convergence of the PC approximation is a function of the variance of the true random variable. To illustrate, consider the lognormal random variable, Y_1 of Table 4.2, with $\alpha = 0$. A decrease in β will decrease the variance of Y_1, allowing an accurate PC approximation, $Y_1^{(p)}$, with fewer terms. This is depicted in Fig. 4.12, which shows plots of the normalized variance and kurtosis of the PC approximation, $Y_1^{(p)}$, for different values of β. Note that $\beta = 0.05, 0.5, 1.0$, correspond to $\mathrm{Var}[Y_1] = 0.00251, 0.365, 4.67$, respectively. The dependence of the the rate of convergence of $Y_1^{(p)}$ on the variance of Y_1 is further illustrated by Fig. 4.13. Plotted is the number of terms, p, required to achieve less than 1% error in variance, skewness, and kurtosis of the PC approximation, $Y_1^{(p)}$. For example, with $\mathrm{Var}[Y_1] = 4.67$ ($\beta = 1$), the variance of $Y_1^{(p)}$ is within 1% of the variance of Y_1 for $p \geq 6$, while the kurtosis of $Y_1^{(p)}$ is within 1% of the kurtosis of Y_1 for $p \geq 13$.

4.1.2 Stationary stochastic processes

Consider a non-Gaussian, $\mathbb{R}^{d'}$-valued, stochastic process, $\mathbf{Y}(t)$, $t \in D$, that can be written as a memoryless transformation of a Gaussian vector process, *i.e.*,

$$\mathbf{Y}(t) = \mathbf{g}\left[\mathbf{S}(t)\right], \tag{4.13}$$

where $\mathbf{g}\colon \mathbb{R}^{k'} \to \mathbb{R}^{d'}$ is a deterministic, measurable mapping, and $\mathbf{S}(t)$ is an $\mathbb{R}^{k'}$-valued zero mean, unit variance, stationary Gaussian stochastic process. It is assumed that

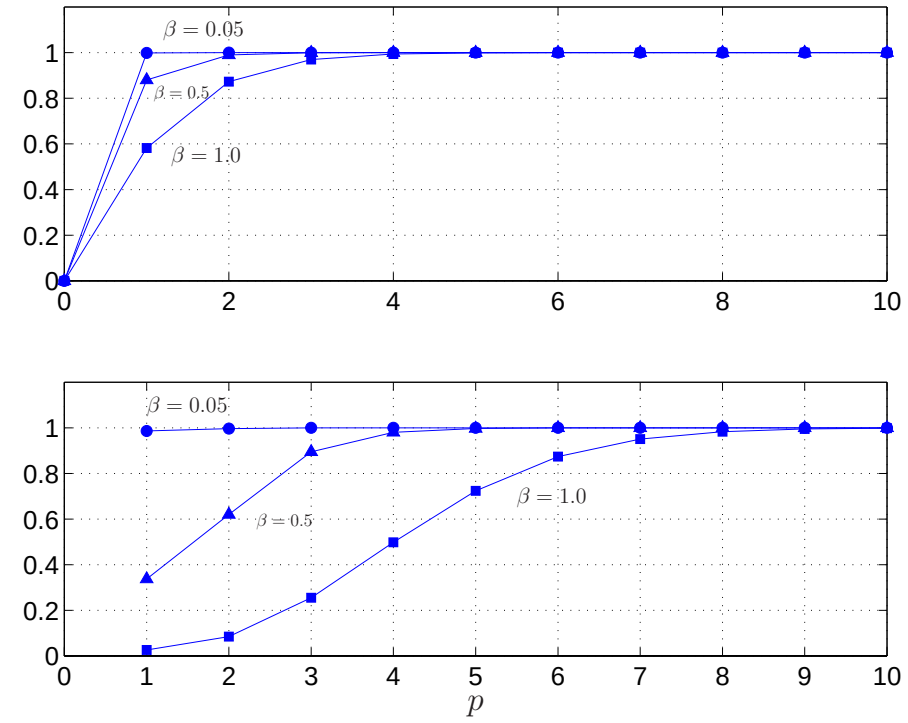

Figure 4.12: Normalized variance (top) and kurtosis (bottom) of $Y_1^{(p)}$.

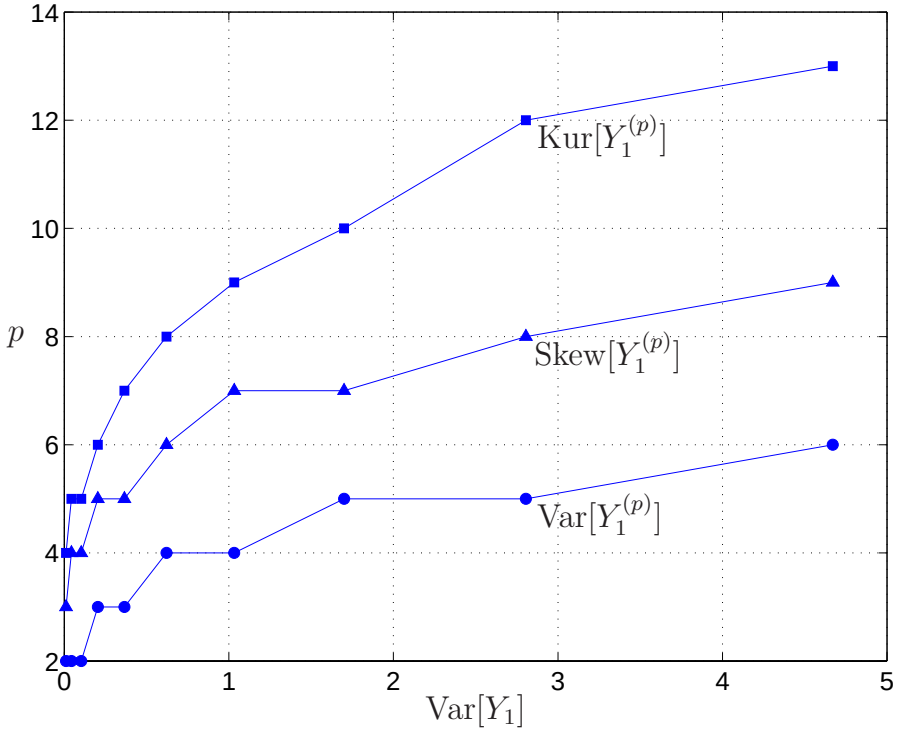

Figure 4.13: Number of terms, p, required to attain 1% error in variance, skewness, and kurtosis of $Y_1^{(p)}$, as a function of $\mathrm{Var}[Y_1]$.

$\mathbf{Y}(t)$ is in L_2, so that all coordinates of $\mathbf{Y}(t)$ have finite second moment. Because $\mathbf{Y}(t)$ is a memoryless transformation of $\mathbf{S}(t)$, it is stationary in the strict sense [48].

4.1.2.1 PC representation for stationary processes

Suppose that the coordinates of $\mathbf{S}(t)$ are independent of each other. Then Eq. (4.2), applied at a fixed time $t \in D$, gives [86, 95]

$$Y_j(t) = g_j\left[\mathbf{S}(t)\right] = \sum_{i_1,i_2,\ldots,i_{k'}\geq 0} y^{(j)}_{i_1,i_2,\ldots,i_{k'}} h_{i_1,i_2,\ldots,i_{k'}}\left[\mathbf{S}(t)\right],\ j = 1,2,\ldots,d', \tag{4.14}$$

where $y^{(j)}_{i_1,i_2,\ldots,i_{k'}}$ are deterministic, time-invariant coefficients and $h_{i_1,i_2,\ldots,i_{k'}}$ denote the orthogonal Hermite polynomials (Eq. (4.3)). Equation (4.14) is the PC representation of the j^{th} coordinate of stationary process $\mathbf{Y}(t)$.

4.1.2.2 PC approximation for stationary processes

Let

$$\begin{aligned} Y^{(p)}_j(t) &= g^{(p)}_j\left[\mathbf{S}(t)\right] \\ &= \sum_{\substack{i_1,i_2,\ldots,i_{k'}\geq 0 \\ i_1+i_2+\cdots+i_{k'}\leq q}} y^{(j)}_{i_1,i_2,\ldots,i_{k'}} h_{i_1,i_2,\ldots,i_{k'}}\left[\mathbf{S}(t)\right], j = 1,2,\ldots,d', \end{aligned} \tag{4.15}$$

be the PC approximation for the random variable $Y_j(t)$, t fixed in D, which consists of Hermite polynomials up to and including order q. The coefficients in Eq. (4.15) can be estimated at any fixed t using orthogonality arguments. Generally, these coefficients must be computed numerically; as previously stated, numerical errors are not considered in this study.

The PC representation and approximation in Eqs. (4.14) and (4.15) for a fixed t coincide with the corresponding formulas for random vectors (Eqs. (4.2) and (4.9)). The extension of Eqs. (4.14) and (4.15) to consider all times $t \in D$ is not direct since the process $\mathbf{S}(t)$ generally consists of an uncountable number of Gaussian random variables. It is therefore necessary to approximate $\mathbf{S}(t)$ by a parametric process, *i.e.*, a finite sum of deterministic functions of time with random coefficients. The following two sections consider the mapping $Y(t) = g[S(t)]$ with $d' = k' = 1$, for the cases in which $S(t)$ is a parametric and an arbitrary Gaussian process, respectively. The errors caused by the PC approximation and the parametric approximation for $S(t)$ can then be examined separately.

4.1.2.2.1 Transformations of parametric processes Consider the class of real-valued stationary Gaussian processes with parametric representation given by

$$S(t) = \sum_{j=1}^{k} f_j(t) W_j, \tag{4.16}$$

where $f_j(t)$ are deterministic functions of time, and W_j are independent $N(0,1)$ random variables. The functions $f_j(t)$ cannot be arbitrary if $S(t)$ is to be stationary [48]. Using Eq. (4.15), the PC approximation for $Y(t) = g[S(t)]$ is then

$$Y^{(p)}(t) = \sum_{i=0}^{q} y_i h_i \left(\sum_{j=1}^{k} f_j(t) W_j \right), \ \forall t \in D. \tag{4.17}$$

This can be rewritten as

$$Y^{(p)}(t) = g^{(p)}(\mathbf{W}, t) = \sum_{\substack{i_1, i_2, \ldots, i_k \geq 0 \\ i_1 + i_2 + \cdots + i_k \leq q}} b_{i_1, i_2, \ldots, i_k}(t) h_{i_1, i_2, \ldots, i_k}(\mathbf{W}), \ \forall t \in D, \tag{4.18}$$

where $\mathbf{W} = [W_1 \ldots W_k]^T$, and p is given by Eq. (4.10). The time-dependent coefficients of Eq. (4.18), $b_{i_1, i_2, \ldots, i_k}(t)$, are defined by collecting like powers of W_j from Eq. (4.17) [86]. This is identical to the PC approximation for random vectors of Eq. (4.9) with t in place of j. Since $Y^{(p)}(t)$ in Eq. (4.17) is a memoryless transformation of a stationary Gaussian process, the PC approximation in Eq. (4.18) is a stationary, non-Gaussian process.

Consider for illustration the process

$$Y(t) = F_Y^{-1} \circ \Phi \left[S(t) \right] = g \left[S(t) \right] = \exp \left[S(t) \right], \tag{4.19}$$

where F_Y is the CDF of the lognormal random variable in the first row of Table 4.2, with ($\alpha = 0$, $\beta = 1$). Further, $S(t)$ is a zero-mean, unit variance, stationary Gaussian process defined as

$$S(t) = \frac{\sqrt{2}}{2} \left[\cos(\pi t) W_1 + \sin(\pi t) W_2 + \cos(2\pi t) W_3 + \sin(2\pi t) W_4 \right], \tag{4.20}$$

so that $k = 4$ in Eq. (4.16). The lognormal process, $Y(t) = \exp[S(t)]$, is strictly stationary, $\mathrm{E}[Y(t)] = \exp(1/2)$, and has a covariance function given by [48]

$$c_Y(\tau) = \exp \left[1 + \frac{\cos(\pi\tau)}{2} + \frac{\cos(2\pi\tau)}{2} \right] - \exp(1). \tag{4.21}$$

Consider the PC approximation for $Y(t)$ in Eq. (4.18) with $q = 2$. Elementary calculations show that $Y^{(p)}(t)$ has mean $\exp(1/2)$ and covariance function

$$c_{Y^{(p)}}(\tau) = y_0^2 + \frac{y_1^2}{2} \left[\cos(\pi\tau) + \cos(2\pi\tau) \right] + \frac{y_2^2}{2} \left[\cos(\pi\tau) + \cos(2\pi\tau) \right]^2, \tag{4.22}$$

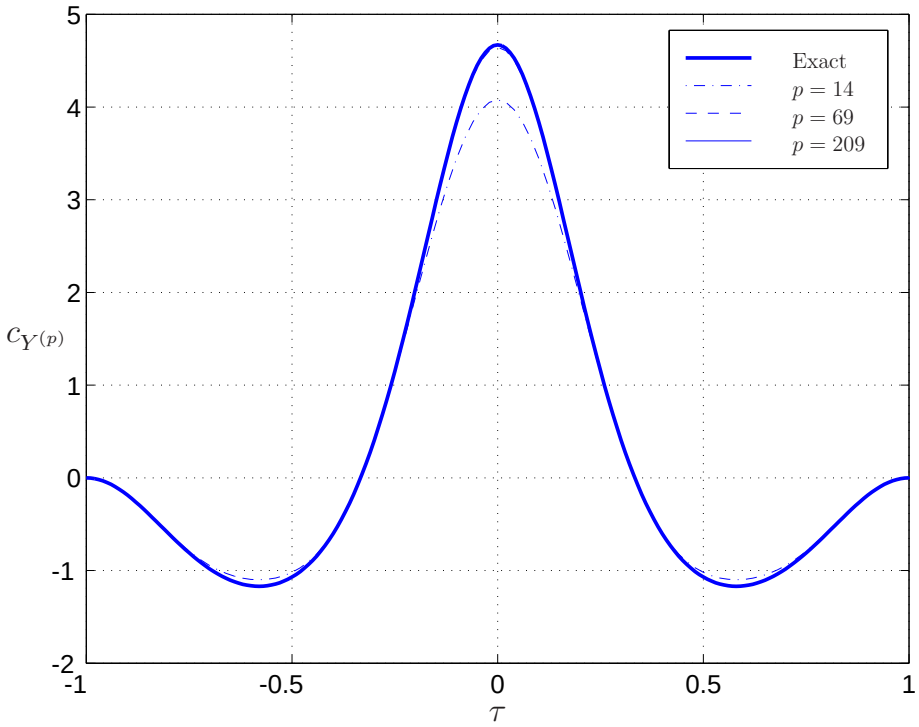

Figure 4.14: Covariance of $Y(t) = g[S(t)]$ and $Y^{(p)}(t)$ for $q = 2, 4, 6$; $S(t)$ is parametric.

where the y_i are the PC coefficients from the lognormal random variable example, $Y_1^{(p)}$, in Table 4.2. The variance of $Y^{(p)}(t)$ approaches $\mathrm{Var}[Y(t)]$ for every t. The covariance function, $c_{Y^{(p)}}(\tau)$, is plotted in Fig. 4.14 for $q = 2, 4, 6$. It shows good agreement for all τ and $q \geq 6$. However, the PC approximation for the stochastic process $Y(t)$ involves more terms than for the lognormal random variable, Y_1, considered previously. By Eq. (4.10), the PC approximations for $Y(t)$ with $q = 2, 4$, and 6 involves $p = 14, 69$, and 209 terms, respectively.

4.1.2.2.2 Transformations of general processes Next let $S(t)$, $t \in [a', a']$, be a general zero-mean, stationary Gaussian stochastic process with covariance function $c_S(\tau)$. The Karhunen-Loève (K-L) representation

$$S(t) = \sum_{j=1}^{\infty} \sqrt{\lambda_j}\phi_j(t)W_j, \tag{4.23}$$

can be used to obtain a parametric approximation for $S(t)$ [43]. In this representation, $\{\lambda_j, \phi_j(t)\}$ are the eigenvalues and eigenfunctions, respectively, of $c_S(\tau)$, and satisfy the integral equation

$$\int_{-a'}^{a'} c_S(t, u)\phi_j(u)\mathrm{d}u = \lambda_j\phi_j(t), \;\; j = 1, 2, \ldots \tag{4.24}$$

The random variables W_j of Eq. (4.23) are given by

$$W_j = \frac{1}{\sqrt{\lambda_j}} \int_{-a'}^{a'} S(t)\phi_j(t)\mathrm{d}t,\ j = 1, 2, \ldots, \tag{4.25}$$

showing that W_j are iid $N(0,1)$ random variables if $S(t)$ is a Gaussian process [42].

In practice, the infinite series of Eq. (4.23) must be truncated, yielding the K-L approximation for $S(t)$, defined as

$$S^{(k)}(t) = \sum_{j=1}^{k} \sqrt{\lambda_j}\phi_j(t)W_j,\ t \in [-a', a']. \tag{4.26}$$

The variance of $S^{(k)}(t)$ satisfies the condition

$$\mathrm{Var}\left[S^{(k)}(t)\right] = \sum_{j=1}^{k}(\sqrt{\lambda_j}\ \phi_j(t))^2 \leq \sum_{j=1}^{\infty}(\sqrt{\lambda_j}\ \phi_j(t))^2 = \mathrm{Var}\left[S(t)\right]. \tag{4.27}$$

By Eq. (4.27) it is evident that: (1) Var$[S(t)]$ is underestimated, and (2) $S^{(k)}(t)$ is not stationary.

There are alternative methods for obtaining parametric approximations for stationary Gaussian processes (see Appendix B). Let $w(\nu)$, having non-zero ordinate in the frequency band $[0, \bar{\nu}]$, $0 < \bar{\nu} < \infty$, denote the one-sided spectral density function of the process $S(t)$ in Eq. (4.23). Consider a partition (u_{j-1}, u_j), $j = 1, 2, \ldots, n$, of $[0, \bar{\nu}]$, where $u_0 = 0$ and $u_n = \bar{\nu}$. Denote by ν_j the mid point of (u_{j-1}, u_j), $j = 1, 2, \ldots, n$, and let

$$\sigma_j^2 = \int_{u_{j-1}}^{u_j} w(\nu)\mathrm{d}\nu. \tag{4.28}$$

The process

$$S^{(n)}(t) = \sum_{j=1}^{n} \sigma_j \left[A_j \cos\left(\nu_j t\right) + B_j \sin\left(\nu_j t\right)\right], \tag{4.29}$$

with $\{A_j, B_j\}$ iid $N(0,1)$, $j = 1, 2, \ldots, n$, (1) provides a parametric approximation for $S(t)$, (2) is stationary, Gaussian with mean zero and covariance function

$$c_{S^{(n)}}(\tau) = \sum_{j=1}^{n} \sigma_j^2 \cos\left(\nu_j \tau\right) \tag{4.30}$$

for any partition of $[0, \bar{\nu}]$ and value of n, and (3) becomes a version (*i.e.*, has an identical finite dimensional distribution) of $S(t)$ as the partition of $[0, \bar{\nu}]$ is refined [49, 50].

To assess the accuracy of the parametric approximations in Eqs. (4.26) and (4.29), consider a Gaussian stochastic process, $S(t)$, $t \in [-a', a']$, with second-moment properties

$$\mathrm{E}\left[S(t)\right] = 0,\ c_S(\tau) = \mathrm{E}\left[S(t)S(t+\tau)\right] = \exp\left(-\mid \tau \mid\right). \tag{4.31}$$

The eigenvalues of the prescribed covariance function are given by [91],

$$\lambda_j = \frac{2}{\theta_j^2 + 1},\ j = 1, 2, \ldots, \tag{4.32}$$

where the θ_j come from the solution to the following characteristic equation,

$$\begin{aligned} 1 - \theta_{2j} \tan\left(a'\theta_{2j}\right) &= 0, \\ \theta_{2j-1} + \tan\left(a'\theta_{2j-1}\right) &= 0. \end{aligned} \tag{4.33}$$

The eigenfunctions are then given by

$$\begin{aligned} \phi_{2j}(t) &= \frac{\cos\left(\theta_{2j}t\right)}{\sqrt{a' + \frac{\sin\left(2a'\theta_{2j}\right)}{2\theta_{2j}}}},\ \text{and} \\ \phi_{2j-1}(t) &= \frac{\sin\left(\theta_{2j-1}t\right)}{\sqrt{a' - \frac{\sin\left(2a'\theta_{2j-1}\right)}{2\theta_{2j-1}}}},\ j = 1, 2, \ldots \end{aligned} \tag{4.34}$$

Using Eqs. (4.32) and (4.34), the K-L approximation in Eq. (4.26) provides a parametric approximation for $S(t)$, $t \in [-a', a']$.

The variance of $S^{(k)}(t)$ is illustrated in Fig. 4.15 for $a' = 5$ and three values of k. As k increases, $\mathrm{Var}[S^{(k)}(t)]$ fluctuates less and approaches one. The rate of convergence is relatively slow, *e.g.*, $\mathrm{Var}[S^{(k)}(t)]$ is in error by 20% for $k = 10$. The scaled K-L approximation [86], $S^{(k)}(t)/\sqrt{\mathrm{Var}[S^{(k)}(t)]}$, has unit variance at all times, $t \in [-a', a']$, but is not a stationary process [33].

Consider now the parametric approximation in Eq. (4.29) with $n = 10, 20, 30$ and $\bar{\nu} = 20$. The covariance function of $S^{(n)}(t)$ is shown in Fig. 4.16 for $n = 10, 20, 30$, along with the covariance of $S(t)$. For $n = 20$, the variance of the approximation, shown at $\tau = 0$, is within 3% of the true variance. All second moment properties of $S^{(n)}(t)$ are accurate for $n \geq 30$.

Regardless of the method used to parameterize $S(t)$, the error introduced by the approximation in Eq. (4.26) or Eq. (4.29) is then propagated to the PC approximation for $Y(t) = g[S(t)]$. To illustrate, consider the process defined in Eq. (4.19), where F_Y is the CDF of the lognormal random variable listed in Table 4.2 with ($\alpha = 0$, $\beta = 1$), and $S(t)$ is a Gaussian process with second moment properties in Eq. (4.31). In this example, the K-L approximation in Eq. (4.26) is used when calculating the PC approximation, $Y^{(p)}(t)$.

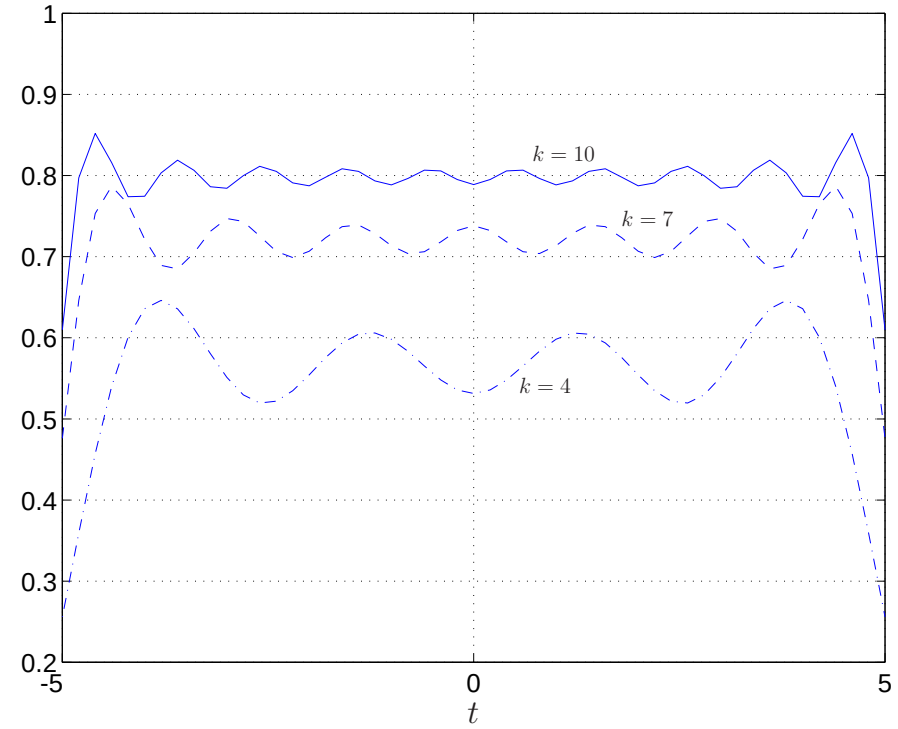

Figure 4.15: Variance of $S^{(k)}(t)$, the K-L approximation of $S(t)$.

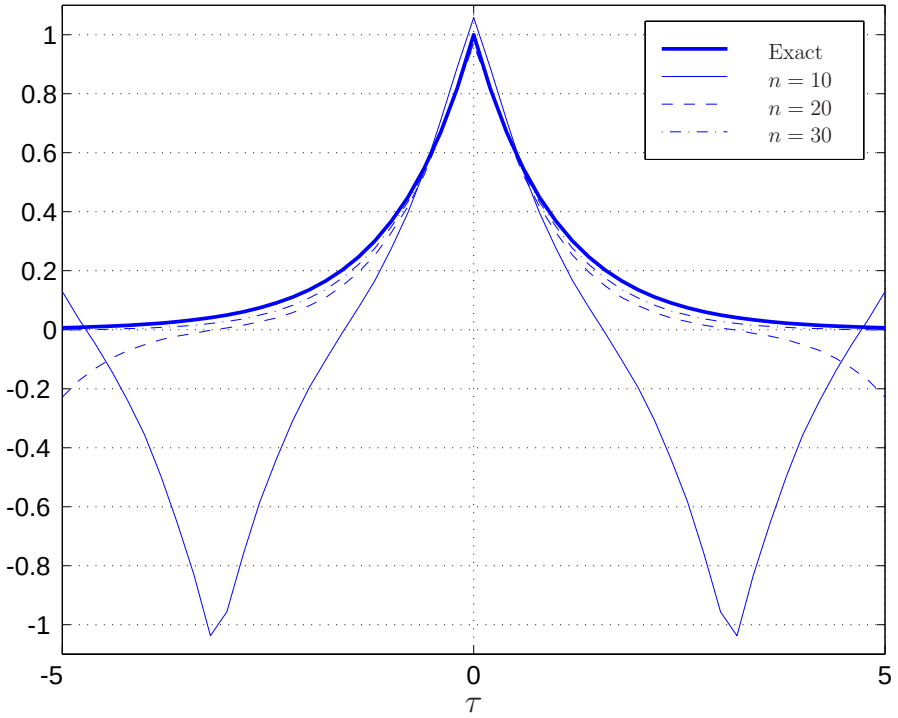

Figure 4.16: Covariance of $S^{(n)}(t)$, the approximation of $S(t)$ by the method of Eq. (4.29).

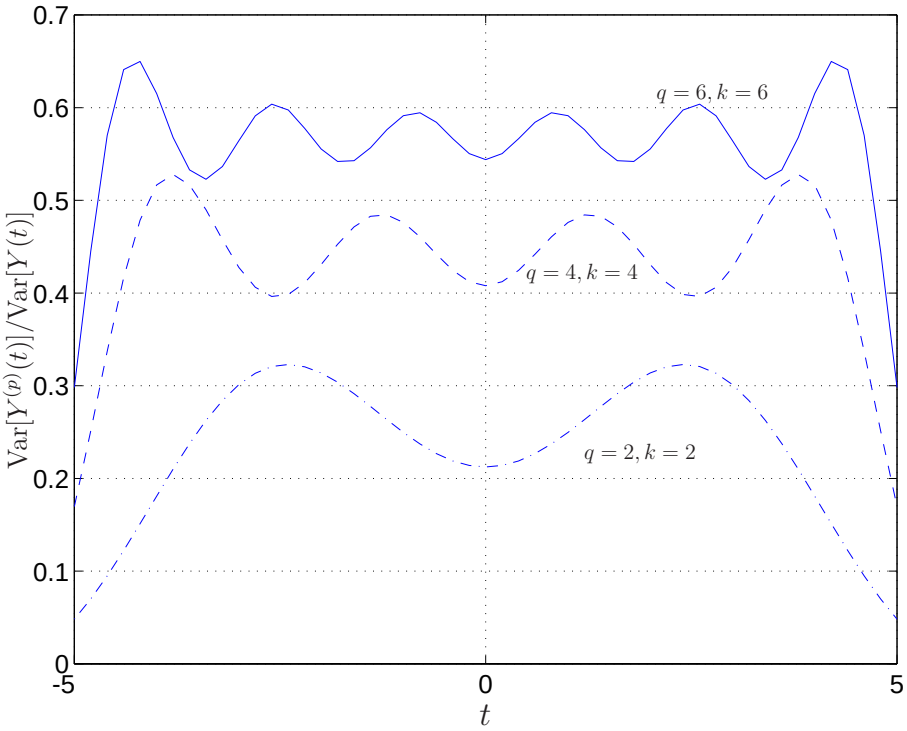

Figure 4.17: Normalized variance of $Y^{(p)}(t)$, with $S(t)$ defined by Eq. (4.31).

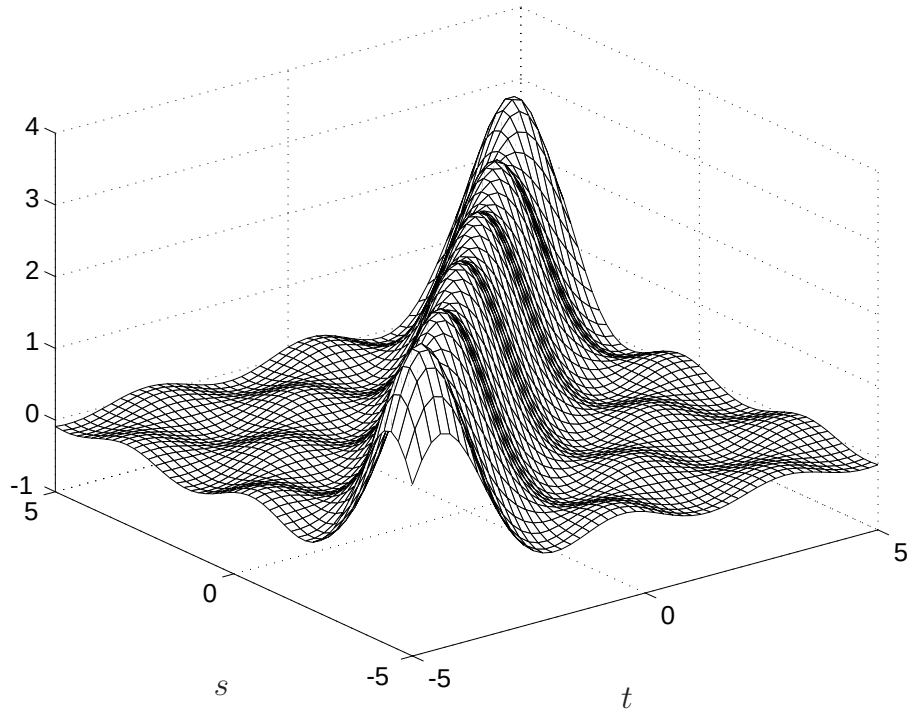

Figure 4.18: Covariance of $Y^{(p)}(t)$ with $(q = 6, k = 6)$ and $S(t)$ defined by Eq. (4.31).

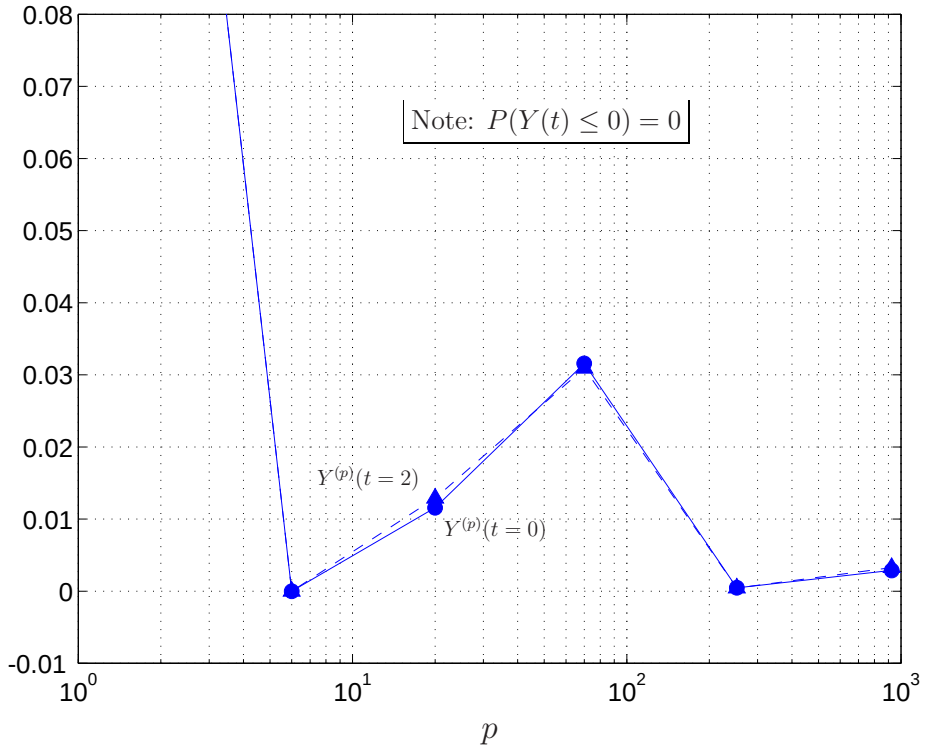

Figure 4.19: Metric #4: estimates of $P(Y_1^{(p)}(t) \leq 0)$.

Figures 4.17 and 4.18 show the scaled variance and the covariance of the PC approximation. As more terms from Eq. (4.18) are retained, $\mathrm{Var}\left[Y^{(p)}(t)\right] / \mathrm{Var}\left[Y(t)\right]$ approaches one as expected. The rate of convergence slows with increasing terms, and even with $(q = 6, k = 6)$, the process remains nonstationary and there is a 40% relative error in the variance of the process.

The probability that $Y^{(p)}(t)$ is negative at $t = 0$ and $t = 2$ is plotted in Fig. 4.19. This probability is high for small p, but becomes nearly zero for large values of p. There is little difference in the two time points. Shown in Fig. 4.20 are the predictions of the 1%- upper fractile at $t = 0$ and $t = 2$. The difference of the marginal distributions at $t = 0$ and $t = 2$ is further evidence of the nonstationarity of $Y^{(p)}(t)$. For both figures, results were obtained using 50,000 samples. These results are very similar to that of the lognormal random variable (Figs. 4.8 and 4.9), except that here p grows exponentially. For $k = q$, where k denotes the number of terms retained in the K-L expansion and q is the order of the polynomial chaos, this relationship is approximately

$$p = \exp\left(1.3256q - 1.0156\right). \tag{4.35}$$

For example, p is $20, 70, 252$, and 924 for $k = q = 3, 4, 5$ and 6, respectively. Hence, adding more terms to the PC approximation requires significant additional computational effort.

The practical use of the polynomial chaos (PC) representation of non-Gaussian

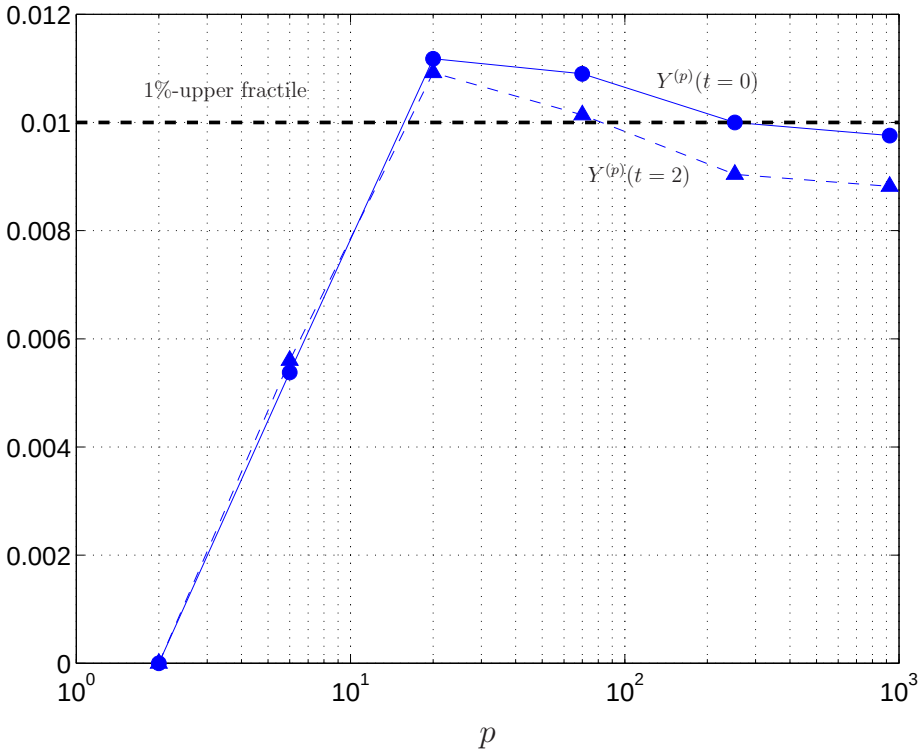

Figure 4.20: Metric #5: estimates of 1%-upper fractile.

random variables and stationary stochastic processes involved several approximations. Each approximation contributed to the total error of the PC approximation. The objective of this section was to identify and quantify some of these errors by simple, relevant example problems. These errors may or may not be significant, depending on the application. In either case, their consequences must be well understood and quantified before using the PC approximation in applications.

The PC approximation converges to the PC representation in mean square for L_2 random variables and stationary stochastic processes. However, this asymptotic property is largely irrelevant for applications since, generally, the PC approximation consists of a small number of terms. We examined the PC approximation for three non-Gaussian random variables and two non-Gaussian, stationary stochastic processes to assess errors of the PC approximation. It was shown that: (1) the accuracy of the PC approximation did not necessarily improve as additional terms were retained, (2) PC approximations for stationary, non-Gaussian stochastic processes may be nonstationary, and (3) the development of PC approximations may be computationally demanding for realistic applications because of the large number of coefficients that need to be calculated.

4.2 Model selection for polynomial chaos

As shown in the previous section, (1) the accuracy of the PC approximation depends on the number of terms, p, retained in the series, and (2) the accuracy does not necessarily improve with increasing p. It is therefore desirable to develop a method to select the optimal order of the PC approximation. In this section, we present an approach to do this that uses the decision-theoretic method for model selection introduced in Chapter 2. The general formulation of the problem is discussed in Section 4.2.1; examples are used to illustrate the method in Section 4.2.2.

4.2.1 General formulation

Recall the non-Gaussian $\mathbb{R}^{d'}$-valued stationary stochastic processes discussed in Section 4.1.2, *i.e.*,

$$\mathbf{Y}(t) = \mathbf{g}[\mathbf{S}(t)],\ t \in D, \tag{4.36}$$

where $\mathbf{g}\colon \mathbb{R}^{k'} \to \mathbb{R}^{d'}$ is a deterministic, measurable mapping, $\mathbf{Y}(t) \in L_2$, and $\mathbf{S}(t)$ is a $\mathbb{R}^{k'}$-valued Gaussian stationary stochastic process with independent coordinates of mean zero and variance one. In the general case, the information on $\mathbf{g}$ is incomplete, so that the probability law for $\mathbf{Y}$ is only partially defined. In this case, we consider a finite collection of possible mappings, given by

$$\mathcal{M} = \{\mathbf{g}_1, \mathbf{g}_2, \ldots, \mathbf{g}_n\}, \tag{4.37}$$

where, for $i = 1, 2, \ldots, n$, each $\mathbf{g}_i\colon \mathbb{R}^{k'} \to \mathbb{R}^{d'}$ in $\mathcal{M}$ is consistent with the available information on $\mathbf{Y}$. Let p_i be the probability that $\mathbf{g}_i$ is true; as noted in Chapter 2, each p_i can be estimated by: (1) the method of maximum likelihood, (2) Bayesian methods, or, (3) if no data is available, noninformative priors.

The objective is to construct the optimal PC approximation for $\mathbf{Y}$. Denote by $g_{i,k}$, $k = 1, 2, \ldots, d'$, the coordinates of $\mathbf{g}_i$. Consistent with Eq. (4.15) from Section 4.1.2.2, denote by $g_{i,k}^{(j)}[\mathbf{S}(t)]$ the j-term PC approximation for $g_{i,k}[\mathbf{S}(t)]$, and let $p \geq 1$ be the maximum order of the PC approximation for $\mathbf{Y}$. The collection

$$\begin{aligned}\mathcal{A} = \Big\{&\mathbf{g}_1^{(1)}[\mathbf{S}(t)], \mathbf{g}_1^{(2)}[\mathbf{S}(t)], \ldots, \mathbf{g}_1^{(p)}[\mathbf{S}(t)], \ldots, \\ &\mathbf{g}_n^{(1)}[\mathbf{S}(t)], \mathbf{g}_n^{(2)}[\mathbf{S}(t)], \ldots, \mathbf{g}_n^{(p)}[\mathbf{S}(t)],\ t \in D\Big\}\end{aligned} \tag{4.38}$$

is the space of candidate actions, where $\mathbf{g}_i^{(j)}[\mathbf{S}(t)] \in \mathcal{A}$ has coordinates $g_{i,k}^{(j)}[\mathbf{S}(t)]$, $k = 1, 2, \ldots, d'$. Each element of $\mathcal{A}$ is a candidate PC approximation for $\mathbf{Y}$; in general, $\mathcal{M}$ and $\mathcal{A}$ do not coincide.

The utility, $U\colon \mathcal{A} \times \mathcal{M} \to [0, \infty)$ and consistent with the definition in Section 2.3.2, is given by

$$U\left(\mathbf{g}_i^{(j)}, \mathbf{g}_k\right) = \gamma\left(\mathbf{g}_i^{(j)}\right) + \psi\left(\mathbf{g}_i^{(j)}, \mathbf{g}_k\right),\ j = 1, \ldots, p, \tag{4.39}$$

for each $i, k \in \{1, 2, \ldots, n\}$, where $\gamma(\mathbf{g}_i^{(j)}) \geq 0$ denotes the cost of model $\mathbf{g}_i^{(j)}$, and $\psi(\mathbf{g}_i^{(j)}, \mathbf{g}_k) \geq 0$ denotes the penalty of model $\mathbf{g}_i^{(j)}$, if $\mathbf{g}_k$ is true. It follows that the expected utility of each candidate PC approximation for $\mathbf{Y}$ is given by Eq. (2.45). The optimal action, $a^\star \in \mathcal{A}$, minimizes Eq. (2.45); we refer to $a^\star$ as the optimal PC approximation for $\mathbf{Y}$.

4.2.2 Examples

Consider the scalar random variable case, *i.e.*, $d = k' = 1$ and $t \in D$ is fixed. Under these constraints, Eq. (4.36) reduces to

$$Y = g(W) = F^{-1} \circ \Phi(W), \tag{4.40}$$

where $W \sim N(0, 1)$ is a standard Gaussian random variable, and F and Φ denote CDFs for Y and W, respectively. Further, $\mathcal{M}$ and $\mathcal{A}$ reduce to

$$\mathcal{M} = \{g_1, g_2, \ldots, g_n\}, \tag{4.41}$$

and

$$\mathcal{A} = \Big\{ g_1^{(1)}(W), g_1^{(2)}(W), \ldots, g_1^{(p)}(W), \ldots, g_n^{(1)}(W), g_n^{(2)}(W), \ldots, g_n^{(p)}(W) \Big\}, \tag{4.42}$$

respectively where, for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, p$, Eq. (4.9) gives

$$g_i^{(j)}(W) = \sum_{l=0}^{j} y_{i,l} h_l(W). \tag{4.43}$$

The coefficients of Eq. (4.43) follow from Eq. (4.8), *i.e.*,

$$y_{i,l} = \frac{1}{\sqrt{2\pi} l!} \int_{-\infty}^{\infty} g_i(u) h_l(u) \exp\left(-u^2/2\right) \mathrm{d}u, \; l = 0, 1, \ldots, j. \tag{4.44}$$

We consider the case where the model use is to predict $P(Y \leq d)$, where d denotes a specified critical value for Y. The objective is then to determine the optimal PC approximation for Y that is consistent with the model use.

The model use is prediction, so the utility formulation of Section 2.3.2.3 applies. Let the cost of model $g_i^{(j)} \in \mathcal{A}$ be given by

$$\gamma\left(g_i^{(j)}\right) = \gamma_0 j, \; i = 1, 2, \ldots, n, \; j = 1, 2, \ldots, p, \tag{4.45}$$

so that PC approximations for Y with a large number of terms have a high cost. Denote by F_k and $F_{i,j}$ the probability laws for $g_k(W) \in \mathcal{M}$ and $g_i^{(j)}(W) \in \mathcal{A}$, respectively, and define

$$\begin{aligned} \psi\left(g_i^{(j)}, g_k\right) &= \tilde{\psi}\left(F_{i,j}(d), F_k(d)\right) \\ &= \begin{cases} \beta_1 \left[F_{i,j}(d) - F_k(d)\right]^2 & \text{if } F_{i,j}(d) \leq F_k(d), \\ \beta_2 \left[F_{i,j}(d) - F_k(d)\right]^2 & \text{if } F_{i,j}(d) > F_k(d), \end{cases} \end{aligned} \tag{4.46}$$

to be the penalty assigned under model $g_i^{(j)} \in \mathcal{A}$, if $g_k \in \mathcal{M}$ is true, where $0 \leq \beta_1 < \beta_2$ are deterministic parameters. By Eq. (4.46), non-conservative predictions of $P(Y \leq d)$ are assigned a penalty; overly conservative predictions are also penalized.

Three cases will be considered in the following sections: (1) the information on Y is complete, (2) the information on Y is limited to a collection of possible probability laws for Y, and (3) only the second-moment properties of Y are known. The decision-theoretic method for model selection will be used in all cases to select the optimal PC approximation for Y.

4.2.2.1 Case #1: complete information on random variable

We assume the probability law for Y, denoted by F, is known and fixed. Hence,

$$\mathcal{M} = \{g\}, \tag{4.47}$$

where $g = F^{-1} \circ \Phi$ is the only possible model for Y, *i.e.*, model g is true with probability one. By Eq. (4.42), we have the following collection of candidate actions

$$\mathcal{A} = \{g^{(1)}(W), g^{(2)}(W), \ldots, g^{(p)}(W)\}, \tag{4.48}$$

where $g^{(i)}$ denotes $g_1^{(i)}$ from Eq. (4.43). By Eq. (4.39), $U(g^{(i)}, g) = \gamma(g^{(i)}) + \psi(g^{(i)}, g)$ is the utility of model $g^{(i)} \in \mathcal{A}$; because the information on Y is complete, the utility is a deterministic function. It follows that

$$u\left(g^{(i)}\right) = \gamma\left(g^{(i)}\right) + \psi\left(g^{(i)}, g\right), \; i = 1, 2, \ldots, p \tag{4.49}$$

is the expected utility of model $g^{(i)} \in \mathcal{A}$.

The expected utility of each PC approximation in $\mathcal{A}$ is computed for the case of Y uniformly distributed in $[0, 2]$, *i.e.*, Example (2) from Table 4.2. Parameters $d = 1.99$, $p = 20$, $\beta_1 = 100$, and $\beta_2 = 500$ were used for calculations, and results from 50,000 Monte Carlo samples were used to estimate $F_i(d)$, $i = 1, 2, \ldots, p$. Results are shown

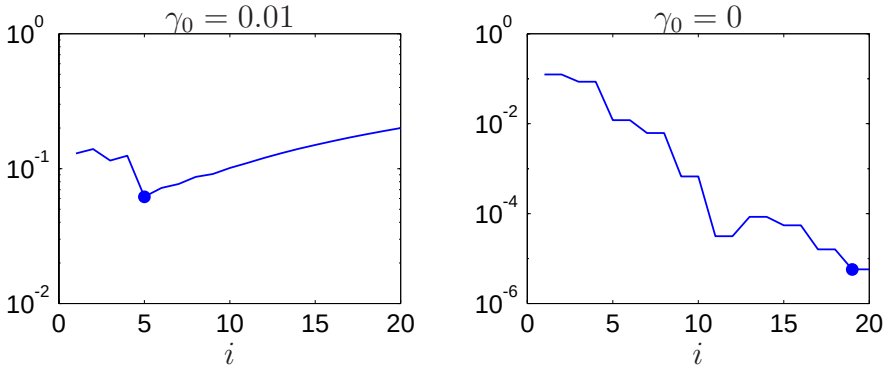

Figure 4.21: Optimal order of PC approximation for case #1.

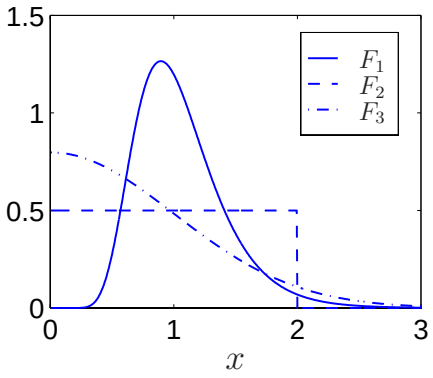

Figure 4.22: PDFs for each $g_k(W) \in \mathcal{M}$ for case #2.

in Fig. 4.21. For the case of $\gamma_0 = 0.01$, shown at the left of Fig. 4.21, the optimal order of the PC approximation for Y is 5, denoted by a solid circle. The plot on the right is for the case of $\gamma_0 = 0$, meaning that we assess zero cost when adding terms to the PC approximation; the optimal order in this case is 19. Hence, for this example, ensuring that the PC approximation does not provide non-conservative predictions of $P(Y \leq d)$ requires that many terms be retained.

4.2.2.2 Case #2: information on random variable limited to a collection of CDFs

We next assume the information on $Y = g(W)$ is limited such that g has one of three possible functional forms, *i.e.*,

$$\mathcal{M} = \{g_1(W), g_2(W), g_3(W)\}, \tag{4.50}$$

where

$$\begin{aligned} g_1(W) &= \exp(\alpha W), \\ g_2(W) &= F_2^{-1} \circ \Phi(W), \text{ and} \\ g_3(W) &= |W|, \end{aligned} \tag{4.51}$$

α is a deterministic parameter, and F_2 is the CDF of a random variable distributed uniformly in $[a, b]$. Note that random variables $g_1(W)$, $g_2(W)$, and $g_3(W)$ are identical to those listed in Table 4.2; the probability laws of each, denoted by F_1, F_2, and F_3, are shown in Fig. 4.22 for the case of $\alpha = 1/3$, $a = 0$, and $b = 2$. No further prior knowledge is available, so each member of $\mathcal{M}$ is assumed equally likely. The collection of PC approximations for the elements of $\mathcal{M}$ defines the space of candidate actions, *i.e.*, Eq. (4.42) with $n = 3$. The utility, $U(g_i^{(j)}, g_k)$, is given by Eq. (4.39), where the cost of $g_i^{(j)} \in \mathcal{A}$ is given by Eq. (4.45), and the penalty associated with $g_i^{(j)} \in \mathcal{A}$ if $g_k \in \mathcal{M}$ is true is given by Eq. (4.46).

The expected utility of each PC approximation in $\mathcal{A}$ is computed by Eq. (2.45); parameters $\gamma_0 = 0$ and $\beta_1 = 100$ were used for calculations. The results from three analyses are listed in Table 4.3. For analyses (1) and (2), $\beta_2/\beta_1 = 1.1$, while for analysis (3), $\beta_2/\beta_1 = 11$, so that non-conservative predictions of $P(Y \leq d)$ are penalized more in analysis (3) than in (1) and (2); analysis (3) corresponds to a case of high risk, as defined in Section 1.3. By Table 4.3, it is evident that the 5-term PC approximation for $g_1(W)$ is the optimal model for analysis (1). If we increase the value for d, the 3-term PC approximation for a lognormal random variable is optimal (analysis (2)). The 16-term PC approximation for $g_3(W)$ is optimal for the high risk case, *i.e.*, analysis (3).

4.2.2.3 Case #3: information on random variable limited to mean and variance

We next assume information on Y is limited to $\mathrm{E}[Y] = \mu < \infty$ and $\mathrm{Var}[Y] = \sigma^2 < \infty$. Under these assumptions, the collection of all possible models for Y is

$$\mathcal{M} = \left\{ g = F^{-1} \circ \Phi \colon \int_{-\infty}^{\infty} s \, \mathrm{d}F(s) = \mu, \int_{-\infty}^{\infty} (s-\mu)^2 \, \mathrm{d}F(s) = \sigma^2 \right\}. \tag{4.52}$$

This is an uncountable collection; consider a finite subcollection

$$\mathcal{M}' = \left\{ g_i^{(p)}(W), \ i = 1, 2, \ldots, n \right\} \subset \mathcal{M}, \tag{4.53}$$

where the coefficients satisfy

$$y_{i,0} = \mu \text{ and } \sum_{j=1}^{p} y_{i,j}^2 j! = \sigma^2. \ i = 1, 2, \ldots, n, \tag{4.54}$$

By Eq. (4.54), $\mathcal{M}'$ is a finite collection of p-term PC approximations for Y, where each approximation has mean μ and variance σ^2. No further information is available, so we assume each $g_i^{(p)} \in \mathcal{M}'$ is equally likely. We will consider the special case where $\mathcal{A} = \mathcal{M}'$.

Table 4.3. Optimal PC approximation for case #2.

Analysis	d	β_2	$a^\star \in \mathcal{A}$
(1)	1.95	110	$g_2^{(5)}$
(2)	1.995	110	$g_1^{(3)}$
(3)	1.999	1100	$g_3^{(16)}$

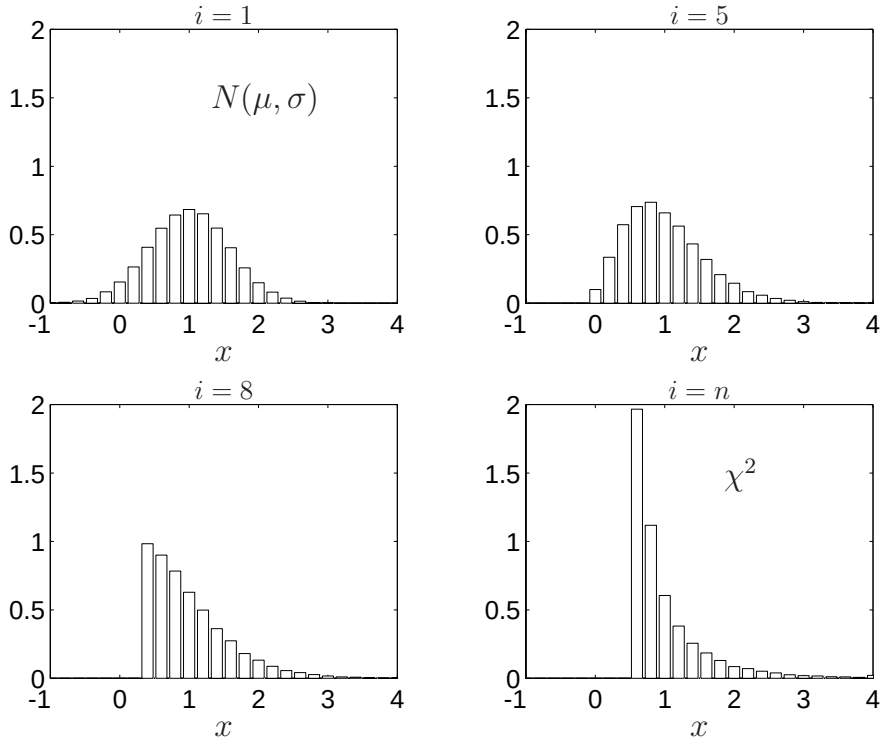

Figure 4.23: Histograms of PC approximations for case #3.

For illustration, let $p = 2$, $n = 21$, and impose the additional constraint

$$y_{i,2} = \frac{\sigma(i-1)}{\sqrt{2}(n-1)},\ i = 1, 2, \ldots, n. \tag{4.55}$$

By Eq. (4.55) the first model in $\mathcal{A}$, $g_1^{(2)}(W) = \mu + \sigma W$, is a Gaussian random variable; likewise, $g_n^{(2)} = \mu + \sigma/\sqrt{2}(W^2 - 1)$ follows a χ^2 distribution with one degree of freedom [5], pp. 172-173. While each $g_i^{(2)} \in \mathcal{A}$ have identical second-moment properties, their probability laws are very different. Histograms of four of the candidate models in $\mathcal{A}$ are shown in Fig. 4.23 for the case of $\mu = 1$ and $\sigma^2 = 1/3$; 50,000 Monte Carlo samples were used.

Because each $g_i^{(2)} \in \mathcal{A}$ has the same number of terms $(p + 1)$, we assign zero cost to every member of $\mathcal{A}$. The utility is given by Eq. (4.46) with $g_i^{(j)}$ and g_k replaced by $g_i^{(2)}$ and $g_k^{(2)}$, respectively. The expected utility of each PC approximation for Y is given by Eq. (2.45); parameters $d = 1.99$ and $\beta_1 = 100$ were used for calculations. The results are shown in Fig. 4.24 for two cases: (1) $\beta_2 = 110$ (left) and (2) $\beta_2 = 1100$ (right). As β_2 increases, the optimal model changes from $g_6^{(2)}$ to $g_{15}^{(2)}$.

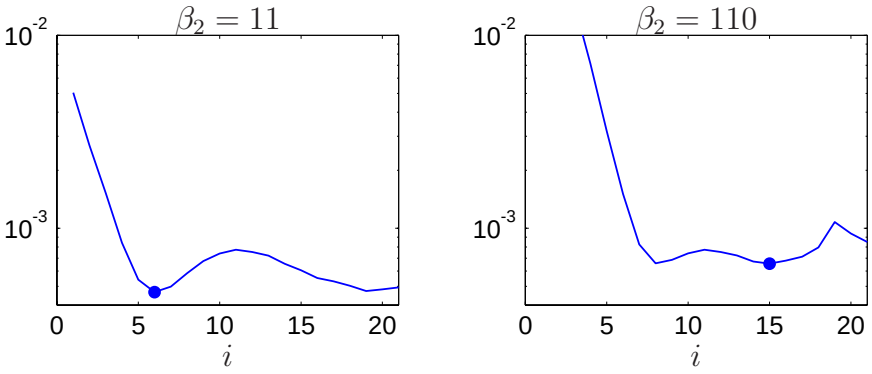

Figure 4.24: Optimal PC approximation for case #3.

Chapter 5

Performance of spacecraft during atmospheric re-entry

Atmospheric re-entry is concerned primarily with the passage of ballistic objects through the Earth's atmosphere [71]. This is one example of a high risk system, as defined in Chapter 1. All space vehicles that are required to return to Earth must endure the extreme environment associated with atmospheric re-entry. This environment includes the effects of aerodynamic heating, large rates of deceleration, radiation effects, and loads due to shock and vibration [30]. While all of these effects must be considered in typical spacecraft design, the work presented herein will be concerned only with the vibrational effects of the re-entry environment; if necessary, the remaining effects can be incorporated into the framework. In this chapter, we use the decision-theoretic method for model selection to select optimal models for the vibration environment to a spacecraft during atmospheric re-entry.

5.1 Motivation

At high speeds encountered during atmospheric re-entry, the in-flight vibration response of a spacecraft is due primarily to external fluctuating pressure loads, which result from unsteady aerodynamic flows over the vehicle [24]. The role of the structural engineer is to ensure that all internal components of the spacecraft, *e.g.*, electronics, astronauts, etc., survive atmospheric re-entry. Typically, this means that the internal components must continue to operate normally with a prescribed level of reliability. To do this requires models of the spacecraft and the re-entry environment, where the latter is a characterization of the applied pressure loads in space and time. Due to the severity of this environment, it is nearly impossible to attain experimental data, and the characteristics of the excitation remain largely undefined. Usually, assumptions are made about the missing information so an analysis can be completed; the consequences of these assumptions are, in general, difficult to quantify.

As discussed in Section 1.3, this can be classified as a high risk system because the consequences of using an inappropriate model for the re-entry environment can be disastrous. The decision-theoretic method for model selection is therefore a suitable technique to select models for the re-entry environment.

In general, information on the applied pressure field during atmospheric re-entry is incomplete. In this chapter, we model the pressure field as a stochastic process in space and time, and examine the effects of two missing pieces of information: (1) the spatial correlation of the process, and (2) the marginal distribution function of the process. The methods for model selection developed in Chapter 2 are used to select the optimal model, from a class of candidate models, for the spatial correlation and marginal distribution of the applied pressure field.

Typically, the assessment of the structural response of a spacecraft during atmospheric re-entry is limited to second-moment analysis [17, 24, 29] and [28], Chapter 10. For structural reliability calculations, this requires one to assume the marginal distribution function of the applied pressure field to be Gaussian. Consider Fig. 5.1, which shows actual accelerometer data recorded during the initial portions of the re-entry phase of mission STS-62 of the NASA Space Shuttle Orbiter [14]. Shown are the accelerations of the center of mass of the Orbiter in three directions as a function of time. The sample functions shown in Fig. 5.1 clearly do not come from a Gaussian process. In addition, wind tunnel pressure measurements on an aircraft fuselage are shown in Fig. 5.2; similar non-Gaussian behavior is noted. Non-Gaussian response (Fig. 5.1) and non-Gaussian applied pressures (Fig. 5.2) together suggest that the applied load during atmospheric re-entry may be non-Gaussian as well.

5.2 Problem description

Consider the spacecraft depicted schematically in Fig. 5.3, where the random vector pressure field, $\mathbf{Z}(\mathbf{x};t)$, models the loading during atmospheric re-entry in space, $\mathbf{x}$, and time, t. The spacecraft, shown here as a perfect cone, has several stiffening ribs oriented in the hoop direction; a base ring is located at the aft end of the spacecraft to provide added stiffness. An internal component, depicted by a solid circle in the figure, is connected to the skin of the spacecraft via several supports, each having structural properties, $\mathbf{k}_c$.

Two response quantities are of interest: the vector displacement response of the skin of the spacecraft, $\mathbf{D}(\mathbf{x};t)$, and the vector acceleration response of the internal component, $\ddot{\mathbf{Y}}(t)$. Models of the spacecraft and the applied pressure field can be developed based on the available information. The objective is to use these models to assess the performance of the spacecraft during atmospheric re-entry from properties of $\mathbf{D}(\mathbf{x};t)$ and $\ddot{\mathbf{Y}}(t)$.

The remainder of the chapter is as follows. In Section 5.3, we develop a model

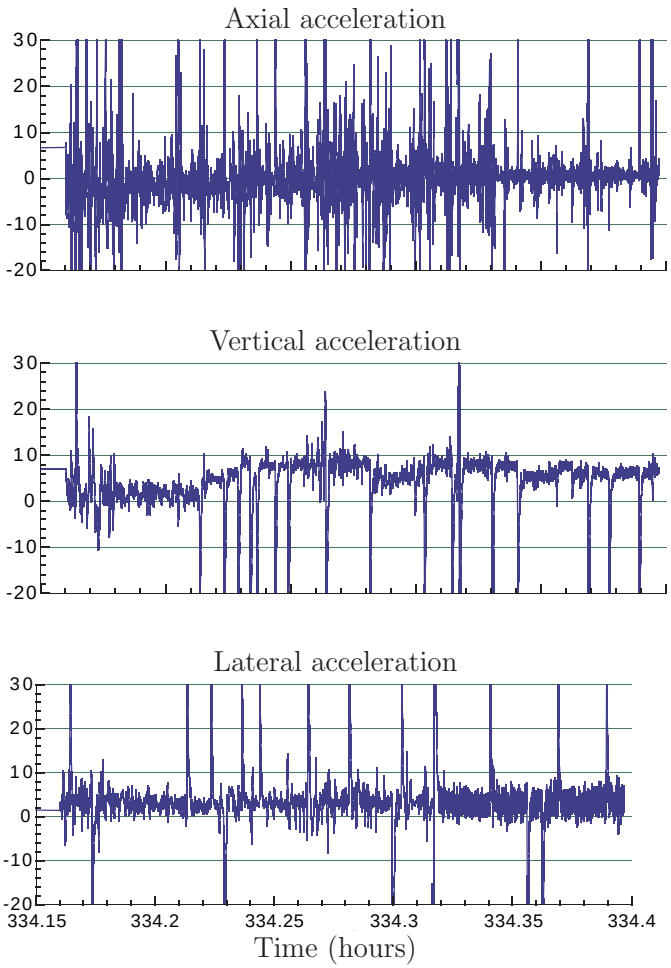

Figure 5.1: Acceleration data from Space Shuttle mission STS-62 [14].

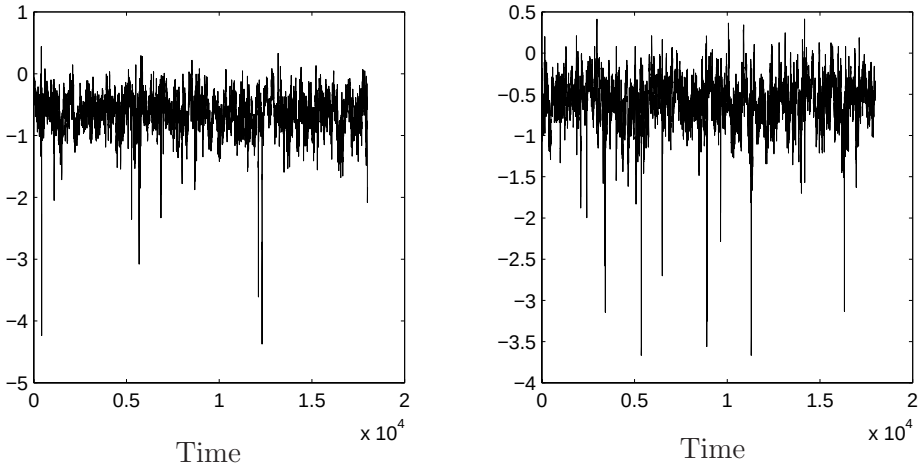

Figure 5.2: Wind tunnel pressure measurements.

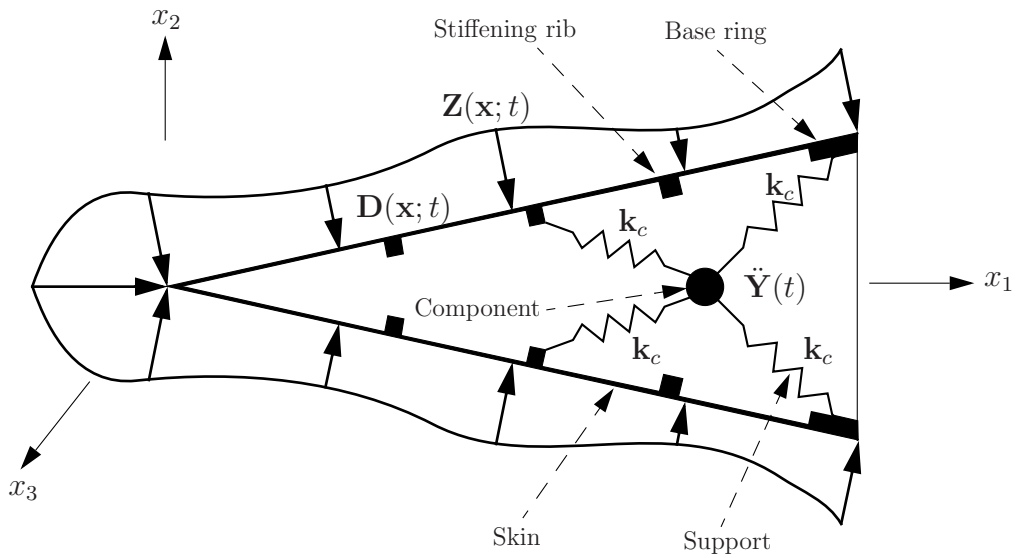

Figure 5.3: Spacecraft during atmospheric re-entry.

for the spacecraft; it is assumed all required information is available to construct the spacecraft model. In Section 5.4, the model for the applied pressure field is developed. Model use is discussed in Section 5.5. The available information on $\mathbf{Z}(\mathbf{x};t)$ is incomplete and we apply methods of model selection to choose optimal models for the input that are consistent with the model use. Because the applied pressure field is a stochastic process in both space and time, the spatial and temporal correlation functions, as well as the marginal distribution function, are necessary to define $\mathbf{Z}(\mathbf{x};t)$. Three cases are considered for illustration: (1) $\mathbf{Z}(\mathbf{x};t)$ is assumed to be a Gaussian process with perfectly known temporal correlation function and spatial correlation function of specified functional form, but with an unknown parameter value, (2) $\mathbf{Z}(\mathbf{x};t)$ is assumed to be a Gaussian process with perfectly known temporal correlation function and spatial correlation function of unspecified functional form with unknown parameters, and (3) $\mathbf{Z}(\mathbf{x};t)$ is assumed to be completely defined in the second-moment sense, but with unknown marginal distribution function. Table 5.1 shows the analysis scenarios to be studied in the remainder of the chapter; the sections corresponding to each scenario are also listed. Methods of model selection are used to select optimal models for $\mathbf{Z}$ under each scenario.

Table 5.1. Analysis scenarios for spacecraft re-entry problem.

Input model	Spacecraft model	
	Case (i)	Case (ii)
Case (1)	Section 5.6	-
Case (2)	Section 5.7	-
Case (3)	-	Section 5.8

5.3 Spacecraft model

In this section, we develop the model for the spacecraft shown in Fig. 5.3. This model defines the operator, $\mathcal{D}$ as defined in Chapter 2, for the re-entry analysis. Two cases are considered: (i) the dynamics of spacecraft and internal component are fully coupled, and (ii) the dynamics of spacecraft and internal component are independent of each other and can be decoupled. Methods for analysis under both case are discussed; cases (i) and (ii) are also shown in Table 5.1. The available information, discussed in Section 5.3.1, is assumed complete. Hence, $\mathcal{D}$ is completely known, fixed, and deterministic. Sections 5.3.2 and 5.3.3 discuss the mathematical model and methods for response analysis, respectively.

5.3.1 Available information

It is assumed that: (1) the spacecraft is made of a single, linear elastic material with known, deterministic modulus of elasticity, e, and mass density, μ, (2) the spacecraft geometry is a perfect cone with cone angle, α, (3) away from the stiffening ribs and base ring, the skin of the spacecraft has constant thickness, c, (4) the component can be modeled as a point mass with value m_c, and (5) its supports behave as simple springs with known, deterministic spring constant, $\mathbf{k}_c$.

Because of the symmetry of the spacecraft, an axial section of the cone, having width $s(x_1) = s_0 x_1$, where $s_0 > 0$ is a constant, can be formulated as a beam on elastic foundation [55]. This conical section is shown in the top of Fig. 5.4; shown at the bottom is α, the cone angle, and c, the skin thickness. The corresponding picture of the spacecraft is a function of a single dimension, $x_1 = x$, and is shown in Fig. 5.5, where $\mathbf{Z}(\mathbf{x};t) = Z(x;t)$ denotes the applied scalar pressure field, and $\mathbf{D}(\mathbf{x};t) = D(x;t)$ denotes the scalar displacement response of the beam. The quantities, Z and D are no longer in bold to denote they are now scalar quantities. The component and its support behave as a single degree-of-freedom oscillator located a distance b from the front of the beam, with mass, m_c, and stiffness, $\mathbf{k}_c = k_c$. The beam has total length l. The quantity $Y(t)$ represents the scalar displacement response of the component with respect to an inertial frame of reference.

5.3.2 Mathematical model

The one-dimensional representation of the spacecraft is a beam on elastic foundation with attached oscillator. The foundation stiffness of the beam is given by (see [55], p. 120)

$$k(x) = \frac{s_0 \; e \; c}{x \; \sin^2 \alpha \; \cos \alpha}, \; 0 < x \leq l, \tag{5.1}$$

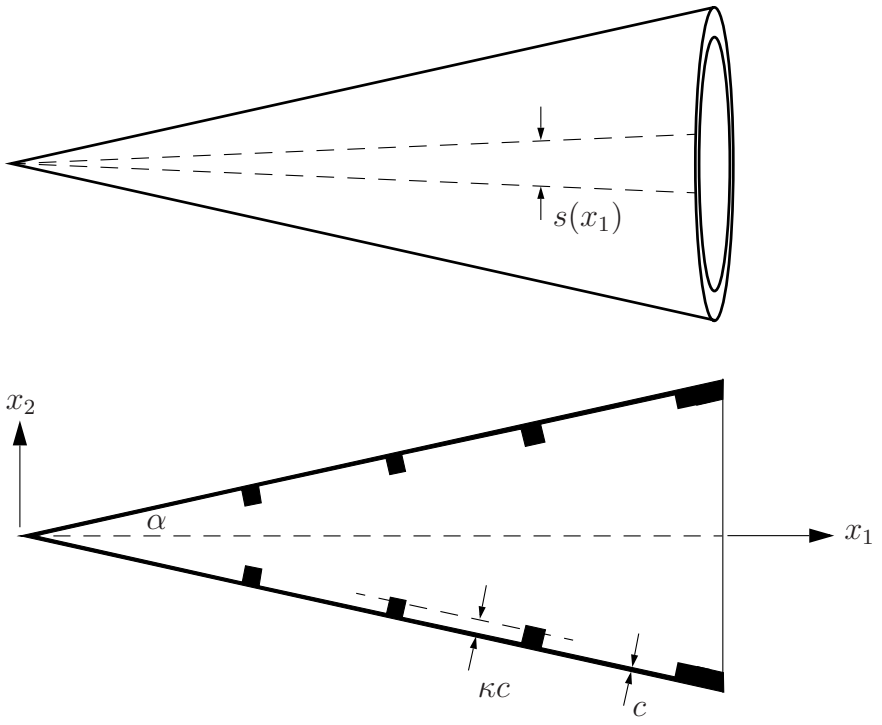

Figure 5.4: Conical section of the spacecraft.

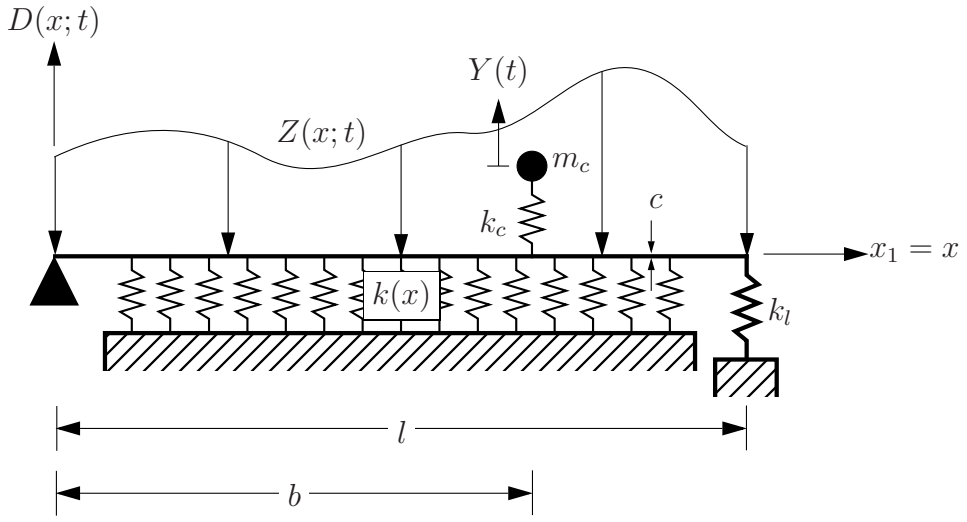

Figure 5.5: One-dimensional spacecraft - beam on elastic foundation with attached oscillator.

Table 5.2. Spacecraft model parameters.

Parameter	Value	Units
α	15	deg
κ	50	-
s_0	0.267	rad/in
μ	0.1	$\mathrm{lb_m/in^3}$
c	0.025	in
e	10.0×10^6	psi
k_l	2.0×10^5	lb/in
l	50	in

where the non-uniformity of $k(x)$ is due to the conical geometry of the spacecraft. At $x = 0$, corresponding to the nose of the spacecraft, the stiffness is infinite; the beam is therefore constrained in the vertical direction at $x = 0$. The beam is tapered, with cross sectional area

$$a(x) = \frac{s_0 \ \kappa \ c}{\cos^3 \alpha} \ x, \tag{5.2}$$

and bending moment of inertia

$$i(x) = \frac{s_0 \ (\kappa \ c)^3}{12 \ \cos^3 \alpha} \ x, \tag{5.3}$$

where parameter $\kappa \geq 1$ is used to represent the effective thickness of the skin due to the stiffening ribs (refer to Fig. 5.4). The spring at the right end of the beam with stiffness $k_l > 0$ is used to represent the added stiffness of the base ring.

Let q and q_c denote viscous damping coefficients for the beam and component support, respectively. The equations of motion governing the beam deflection and position of the mass are given by [10]

$$\begin{aligned} e\left[i(x) \ D''(x;t)\right]'' + \mu a(x)\ddot{D}(x;t) + q\dot{D}(x;t) - k(x)D(x;t) - k_l D(l;t)\delta(x-l) = \\ - m_c\ddot{Y}(t)\delta(x-b) - \frac{xs(x)}{2}Z(x;t), \\ m_c\ddot{Y}(t) + q_c\dot{Y}(t) + k_c Y(t) = q_c\dot{D}(b;t) + k_c D(b;t), \end{aligned} \tag{5.4}$$

with appropriate boundary and initial conditions. Here, $\delta(x)$ denotes the Dirac delta function, $xs(x)/2$ is the area of the top of the beam, used to convert the applied pressure, $Z(x;t)$, to a force, and $(\ )'$ and $\dot{(\ )}$ denote differentiation with respect to x and t, respectively.

Table 5.2 lists the values of the parameters of the spacecraft model. Parameters e and μ are consistent with the properties of aluminum. The base ring stiffness is five

times the foundation stiffness at the right end of the beam, *i.e.*, $k_l = 5k(l)$, and s_0 was selected so that $s(l) = 1$ radian. Parameters c and κ were selected to give adequate separation between the stiffness of the beam and the stiffness of the foundation.

5.3.3 Response analysis

Equation (5.4) cannot be solved for $D(x;t)$ or $\ddot{Y}(t)$ in closed-form, even for the case of deterministic excitation. An alternative is to discretize the beam in the spatial dimension. Define $0 = x_1 < x_2 < \cdots < x_n = l$ to be a partition of $[0, l]$ and let

$$\begin{aligned} D_i(t) &= D(x_i;t),\ i = 1, 2, \ldots, n, \\ \mathbf{D}(t) &= (D_1(t)\ D_2(t)\ \ldots\ D_n(t))^T \end{aligned} \tag{5.5}$$

An approximation to Eq. (5.4) is then given by

$$\mathbf{m}\begin{pmatrix}\ddot{\mathbf{D}}(t)\\ \ddot{Y}(t)\end{pmatrix} + \mathbf{q}\begin{pmatrix}\dot{\mathbf{D}}(t)\\ \dot{Y}(t)\end{pmatrix} + \mathbf{k}\begin{pmatrix}\mathbf{D}(t)\\ Y(t)\end{pmatrix} = \mathbf{b}\mathbf{Z}(t), \tag{5.6}$$

where $\mathbf{m}$, $\mathbf{q}$, and $\mathbf{k}$ are $(n+1)\times(n+1)$ matrices representing the mass, damping, and stiffness of the beam with attached oscillator, respectively, $\mathbf{b}$ is a $(n+1)\times n'$ matrix, and

$$\begin{aligned} Z_i(t) &= Z(x_i;t),\ i = 1, 2, \ldots, n', \\ \mathbf{Z}(t) &= (Z_1(t)\ Z_2(t)\ \ldots\ Z_{n'}(t))^T, \end{aligned} \tag{5.7}$$

is the input, $Z(x;t)$, discretized in the spatial dimension via partition $0 = x_1 < x_2 < \cdots < x_{n'} = l$.

Figure 5.6 shows the first nine mode shapes, $\mathbf{w}_i$, and the corresponding modal frequencies, ν_i, $i = 1, 2, \ldots, 9$, of the one-dimensional model of the spacecraft, *i.e.*, the solution to

$$\mathbf{k}\mathbf{w} = \mathbf{m}\mathbf{w}\boldsymbol{\nu}, \tag{5.8}$$

where $\mathbf{w} = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_{n+1})$, and $\boldsymbol{\nu}$ is a diagonal matrix such that $\boldsymbol{\nu}_{ii} = \nu_i$, $i = 1, 2, \ldots, n+1$. Parameters $b = 37.5$ in, $m_c = 24$ lb$_\text{m}$, and $k_c = 2.5\times10^5$ lb/in were used for calculations. In each plot, the solid circle represents the displacement of the mass. The system is assumed classically damped with constant damping ratio, $\zeta = 0.02$, *i.e.*, $\mathbf{q} = \mathbf{w}^{-T}\tilde{\mathbf{q}}\mathbf{w}^{-1}$, where the $\mathbf{w}_i$ are assumed mass normalized and $\tilde{q}_{ii} = 2\zeta\nu_i$, $i = 1, 2, \ldots, n+1$ [25].

If the mass and the beam can be assumed to be uncoupled, Eq. (5.6) can be simplified to

$$\mathbf{m}'\ddot{\mathbf{D}}(t) + \mathbf{q}'\dot{\mathbf{D}}(t) + \mathbf{k}'\mathbf{D}(t) = \mathbf{b}'\mathbf{Z}(t), \tag{5.9a}$$

$$\ddot{Y}(t) + 2\zeta_c\omega_c\dot{Y}(t) + \omega_c^2 Y(t) = 2\zeta_c\omega_c\dot{D}_j(t) + \omega_c^2 D_j(t), \tag{5.9b}$$

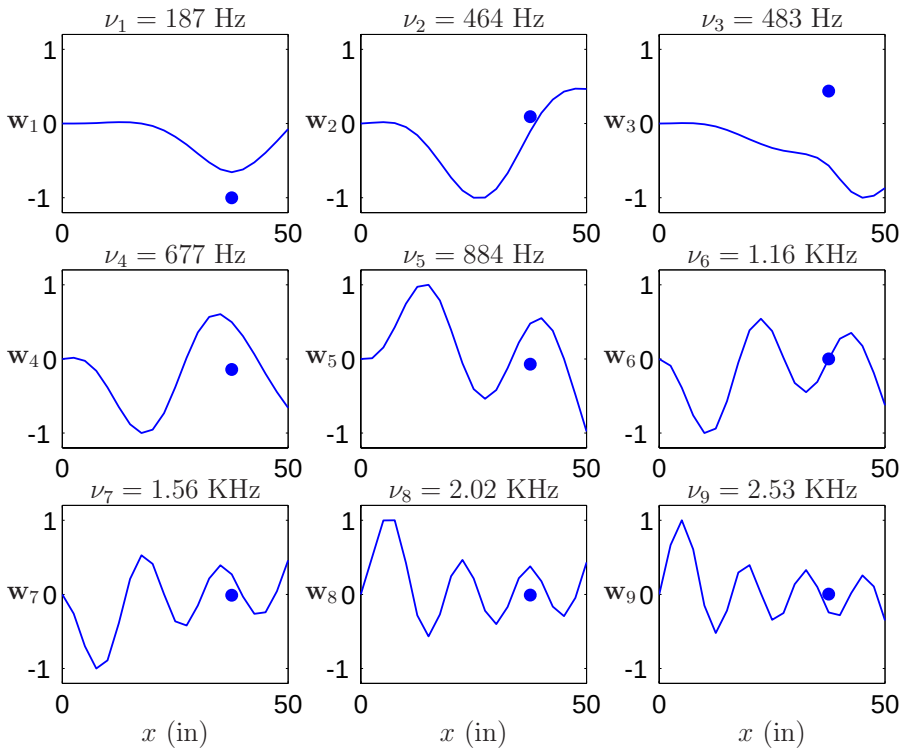

Figure 5.6: Natural frequencies, ν_i, and mode shapes, $\mathbf{w}_i$, $i = 1, 2, \ldots, 9$, of beam on elastic foundation with attached oscillator.

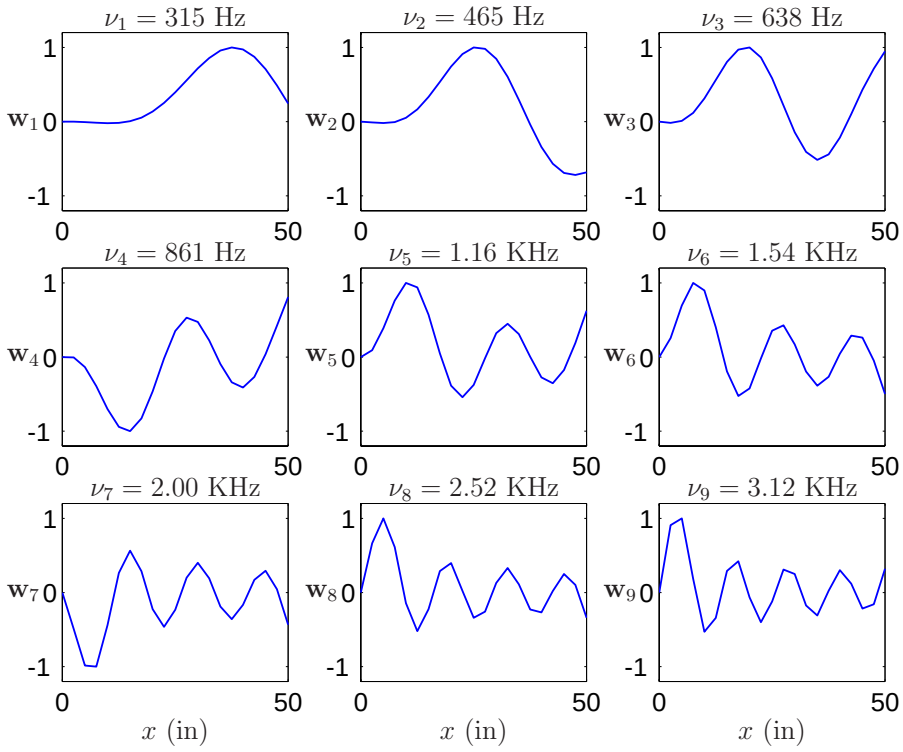

Figure 5.7: Natural frequencies, ν_i, and mode shapes, $\mathbf{w}_i$, $i = 1, 2, \ldots, 9$, of beam on elastic foundation without the oscillator.

where $\mathbf{m}'$, $\mathbf{q}'$, and $\mathbf{k}'$ are $\mathbf{m}$, $\mathbf{q}$, and $\mathbf{k}$, respectively, with the $(n+1)$ row and column removed, $\mathbf{b}'$ is $\mathbf{b}$ with the $(n+1)$ row removed, and $x_j = b$. Parameters $\zeta_c = q_c/\sqrt{4k_c m_c}$ and $\omega_c = \sqrt{m_c/k_c}$ are the damping ratio and natural frequency, respectively, of the oscillator. The natural frequencies and mode shapes of the beam without the oscillator, the eigensolution to Eq. (5.9a), are shown in Fig. 5.7. Equations (5.6) and (5.9) can both be converted into a system of $2(n+1)$ linear ordinary differential equations and solved using a state-space approach. In the analyses that follow, $n = n' = 20$.

5.4 Input model

In this section, we develop the model of the input to the spacecraft, $Z(x;t)$. As will be shown in Section 5.4.1, the available information on Z is incomplete. Hence, the model for the input is not completely defined.

5.4.1 Available information

It is assumed that the applied pressure field can be written as

$$Z(x;t) = \sigma_Z(x)Q(x;t), \ t \in [0,t'], \tag{5.10}$$

where $\sigma_Z(x)$ is a known, non-negative, deterministic function of x shown on the right of Fig. 5.8, $Q(x;t)$ is a weakly stationary stochastic process with partially defined second-moment properties, and $t' \geq 0$ denotes the duration of the re-entry event. By Eq. (5.10), the properties of Q imply that: (1) Z has zero mean (2) the variance of Z is $\sigma_Z^2(x)$, and (3) the temporal and spatial correlation functions of Z are

$$\mathrm{E}[Z(x_0;t)Z(x_0;t+\tau)] = \sigma_Z(x_0)\mathrm{E}[Q(x_0;t)Q(x_0;t+\tau)], \tag{5.11}$$

and

$$\mathrm{E}[Z(x;t_0)Z(x+\eta;t_0)] = \sigma_Z(x)\sigma_Z(x+\eta)\mathrm{E}[Q(x;t_0)Q(x+\eta;t_0)], \tag{5.12}$$

respectively, for any fixed $x_0 \in [0,l]$, and any fixed $t_0 \in [0,t']$.

The available information on Q is given by:

1. $\mathrm{E}[Q(x;t)] = 0$, $\forall x \in [0,l]$, $\forall t \in [0,t']$,
2. $\mathrm{Var}[Q(x;t)] = 1$, $\forall x \in [0,l]$, $\forall t \in [0,t']$, and
3. the temporal correlation function of Q is given by

$$\mathrm{E}[Q(x_0;t)Q(x_0;t+\tau)] = c_Q(\tau) = \begin{cases} \dfrac{\sin(\bar{\omega}\tau)}{\bar{\omega}\tau} & \tau \neq 0, \\ 1 & \tau = 0, \end{cases} \tag{5.13}$$

for any fixed $x_0 \in [0, l]$, where $\bar{\omega} > 0$ is a deterministic parameter, so that Q is a band limited white noise process [88] with cut-off frequency $\bar{\omega}$.

The Fourier transform of Eq. (5.13) gives the power spectral density (PSD) of Q; the one-sided PSD, $g_Q(\omega)$, is shown on the left of Fig. 5.8 for $\bar{\omega} = 20{,}000$ Hz. The properties for Z and Q are consistent with results from both empirical studies [66] and theoretical models [53]. Information on the spatial correlation function of Q,

$$\mathrm{E}[Q(x; t_0)Q(x + \eta; t_0)] = \phi(\eta; \theta), \tag{5.14}$$

for any fixed $t_0 \in [0, t']$, is unavailable. Hence, the spatial correlation function, ϕ, depending on spatial lag, η, and unknown parameter, θ, has unknown functional form.

5.4.2 Mathematical model

As illustrated in the previous section, the second-moment properties of the input, Z, are not completely defined. Information on the spatial correlation of Q and, hence, the spatial correlation of Z, is not available. In addition, nothing is known about the probability laws of Q or Z.

We consider three cases for the missing information on Z, and apply the decision-theoretic method for model selection under each. **Case #1**: $Q = G$, where G is a Gaussian process with spatial correlation function given by

$$\phi(\eta; \theta) = e^{-\theta|\eta|}, \tag{5.15}$$

and $\theta \geq 0$. Here, the functional form for ϕ is known, but the value for θ is not. **Case #2**: $Q = G$, where G is a Gaussian process with spatial correlation function $\phi(\eta; \theta)$ of unknown functional form; the value for the parameter θ is also unknown. **Case #3**: Q is a non-Gaussian process with completely defined second-moment properties.

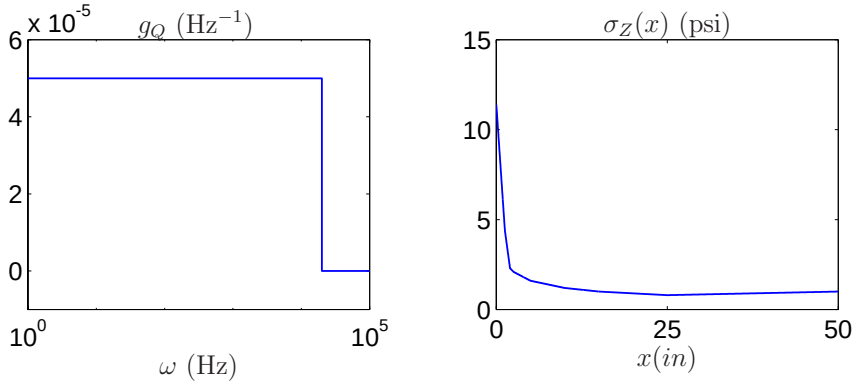

Figure 5.8: Available information on Z.

In this case, we set $\phi(\eta) = 1$, $\forall \eta \in [0, l]$. This corresponds to the limiting case of $\theta \to 0$ in Eq. (5.15).

In sections 5.6, 5.7, and 5.8, we employ the model selection techniques developed in Chapter 2 to choose optimal models for cases #1, #2, and #3, respectively. In each section, we first define a class of candidate models, where each member of the class is consistent with all available information on Z. Second, we illustrate that different models from the class of candidate models may have a significant impact on the properties of the response of interest. Third, we develop an appropriate utility function and use it to select the optimal model from the class of candidate models. In addition, for case #1, we study the sensitivity of the optimal model to changes in the utility function.

5.5 Model use

The objective is to use the mathematical models for the spacecraft and the applied pressure field to assess performance. In particular, suppose

$$g(Y) = \max_{t \in [0, 3t']} | \ddot{Y}(t) | \tag{5.16}$$

is the property of interest, where $\ddot{Y}(t) = d^2Y(t)/dt^2$ and $t' \geq 0$ denotes the duration of the re-entry event. The performance of the system is then assessed via predictions of

$$P\left(g(Y) \leq d\right), \tag{5.17}$$

where d denotes a critical value of $\ddot{Y}$. This is a prediction problem, so the utility formulation of Section 2.3.2.3 applies.

5.6 Case #1: model selection for spatial correlation parameter

In this section, we employ the model selection techniques to select the optimal value for parameter θ of the spatial correlation function of the applied pressure field, Z, defined by Eq. (5.15).

5.6.1 Candidate models

Let

$$\mathcal{M} = \{\theta \geq 0 \colon \phi(\eta; \theta) = \exp\left(-\theta \mid \eta \mid\right)\}. \tag{5.18}$$

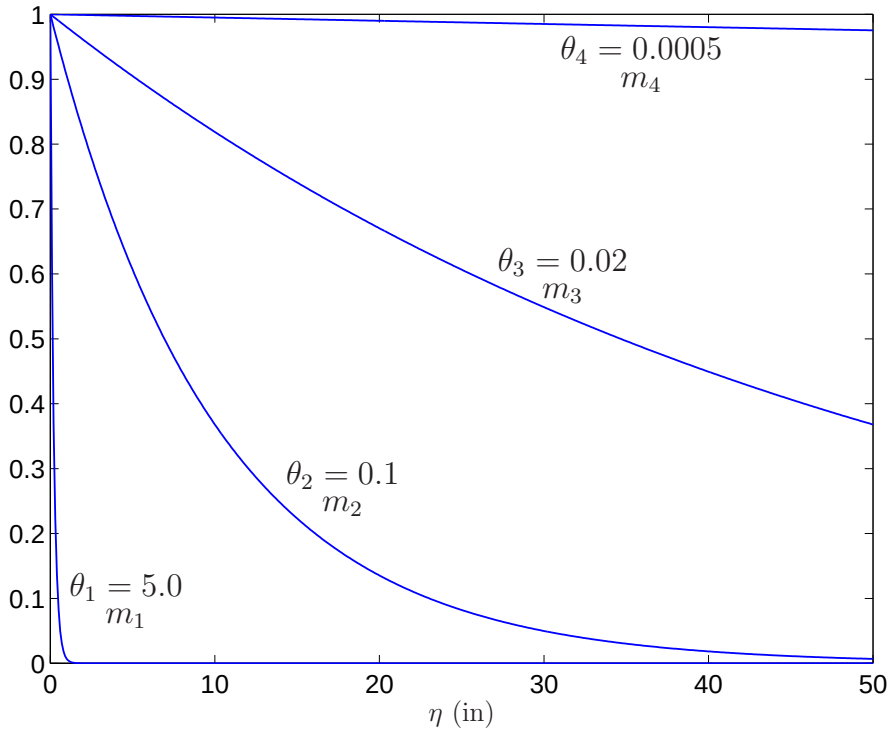

Figure 5.9: Candidate spatial correlation models, $\phi(\eta;\theta_i) = \exp(-\theta_i|\eta|)$.

be a collection of candidate models. This collection is uncountably infinite; we instead consider a finite subcollection $\mathcal{M}' \subset \mathcal{M}$, where

$$\mathcal{M}' = \{m_i(\theta_i)\} = \{\theta_i \geq 0 \colon \phi(\eta;\theta_i) = \exp(-\theta_i \mid \eta \mid), i = 1, 2, \ldots, 4\}. \tag{5.19}$$

The four candidate models are shown in Fig. 5.9. By Eq. (5.15), for large θ $G(x;t)$ is nearly uncorrelated in the spatial dimension; as θ approaches zero, the process becomes perfectly correlated in space. Note that the correlation structure in time for each member of $\mathcal{M}'$, *i.e.*, $\mathrm{E}[G(x_0;t)G(x_0;t+\tau)]$ for fixed $x_0 \in [0,l]$, is equal to that of Eq. (5.13).

One sample of $G(x;t)$ is shown in Fig. 5.10 for each of the four values for θ considered and $t' = 0.03$ sec; methods from Appendix B were used to generate the samples, where $c_{kl}(\tau)$ of Eq. (B.1) is replaced by $c_Q(\tau)\phi(x_k - x_l;\theta)$. Each plot shows $G(x;t)$ at $x = 10$ in (solid line) and $x = 37.5$ in (dashed line), the latter being the attachment point for the oscillator. Note that as θ approaches zero, the processes $G(x_1;t)$ and $G(x_2;t)$, $x_1, x_2 \in [0,l]$, become scaled versions of each other [50]. The corresponding samples of input $Z(x;t)$ via Eq. (5.10) are shown in Fig. 5.11. The magnitude of $Z(x = 10 \text{ in};t)$ is larger than that of $Z(x = 37.5 \text{ in};t)$ because of the scaling by $\sigma_Z(x)$ (Fig. 5.8).

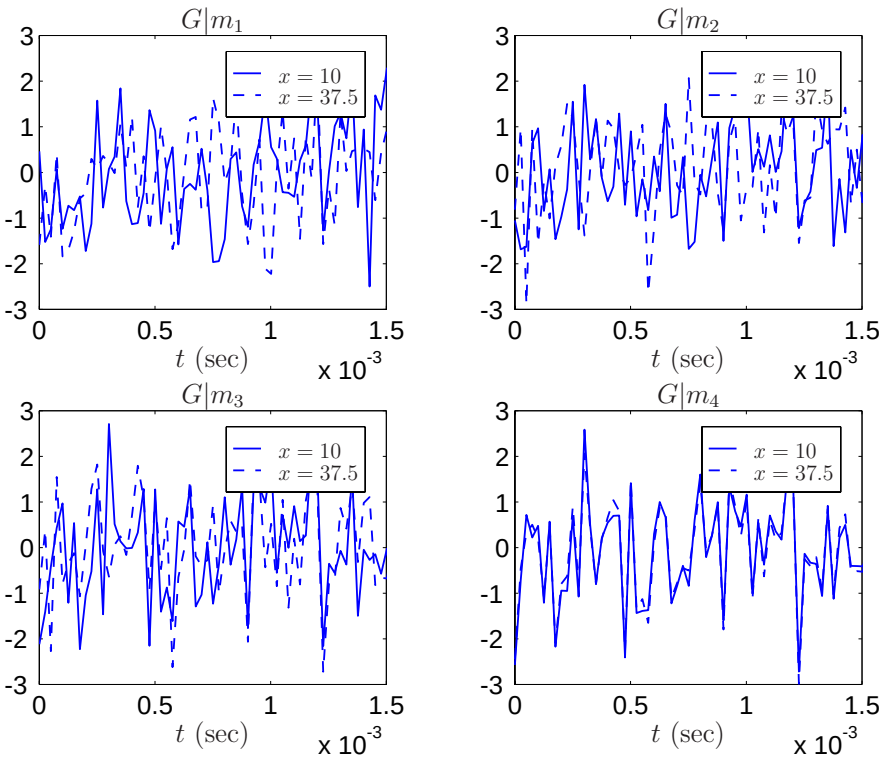

Figure 5.10: One sample of Gaussian process, $G(x;t)$.

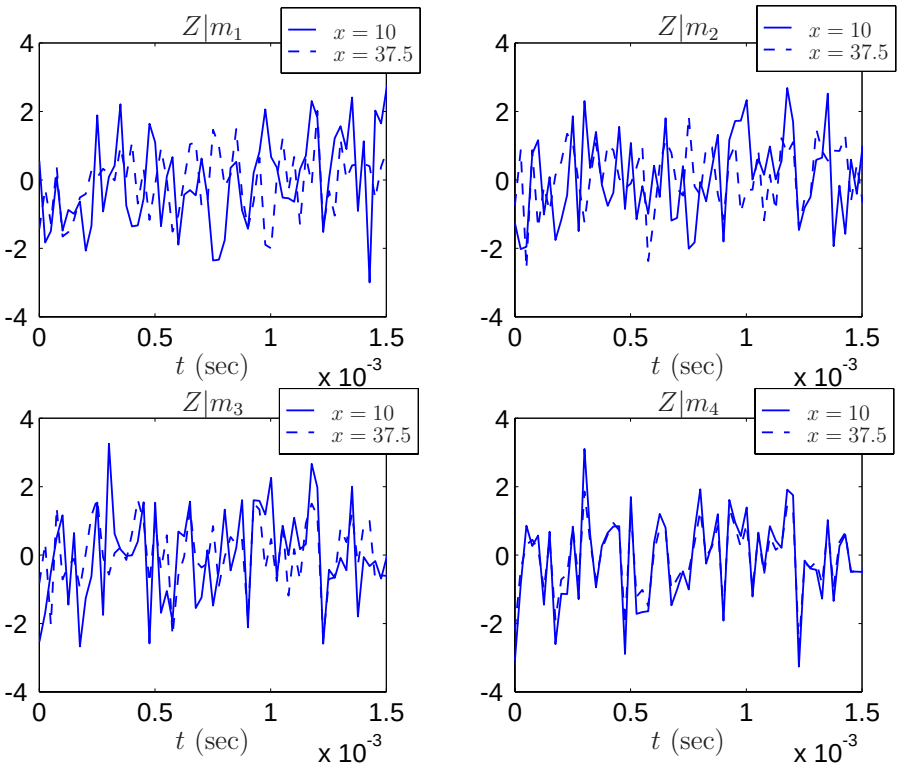

Figure 5.11: One sample of beam input process, $Z(x;t)$, in units of psi.

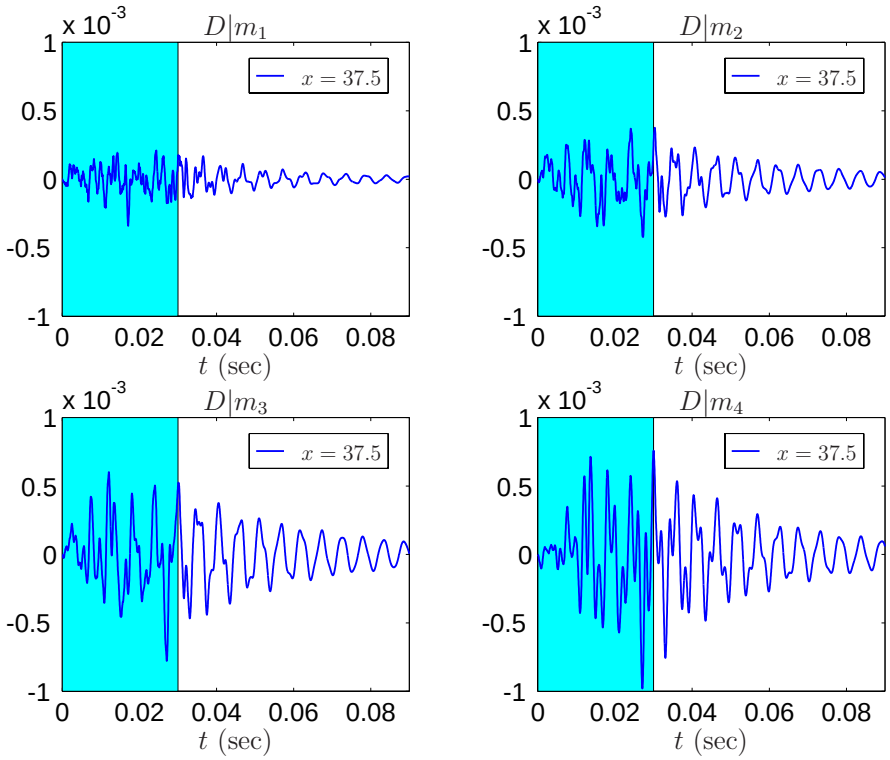

Figure 5.12: One sample of beam response process, $D(x;t)$, in units of in.

5.6.2 Sensitivity of model output

One sample of the displacement response of the beam at the attachment point, $D(x = 37.5 \text{ in}; t)$, is shown in Fig. 5.12, for $t \in [0, 3t']$. The shaded area indicates the time interval when the load is applied, $t \in [0, t']$; the unshaded area corresponds to free vibration. It is evident that the displacement response at the attachment point is very sensitive to the degree of spatial correlation on the input: as the spatial correlation increases (decreasing θ), the magnitude of the response increases.

Estimates of $(\max_x |D(x;t)|, \max_t |D(x;t)|)$, $x \in [0, l]$, $t \in [0, 3t']$, the joint distribution of the location and magnitude of the maximum displacement of the beam, using results from 2,000 Monte Carlo samples, are shown in Fig. 5.13. For large θ (nearly uncorrelated in space), the range of observed values for $\max_x |D(x;t)|$ is large and they occur most frequently at the right end of the beam; the range of observed values for $\max_t |D(x;t)|$ is small in comparison. As θ decreases, the pattern is reversed. The range of observed values for $\max_x |D(x;t)|$ is small and they occur most frequently near the attachment point; the range of observed values for $\max_t |D(x;t)|$ is large in comparison.

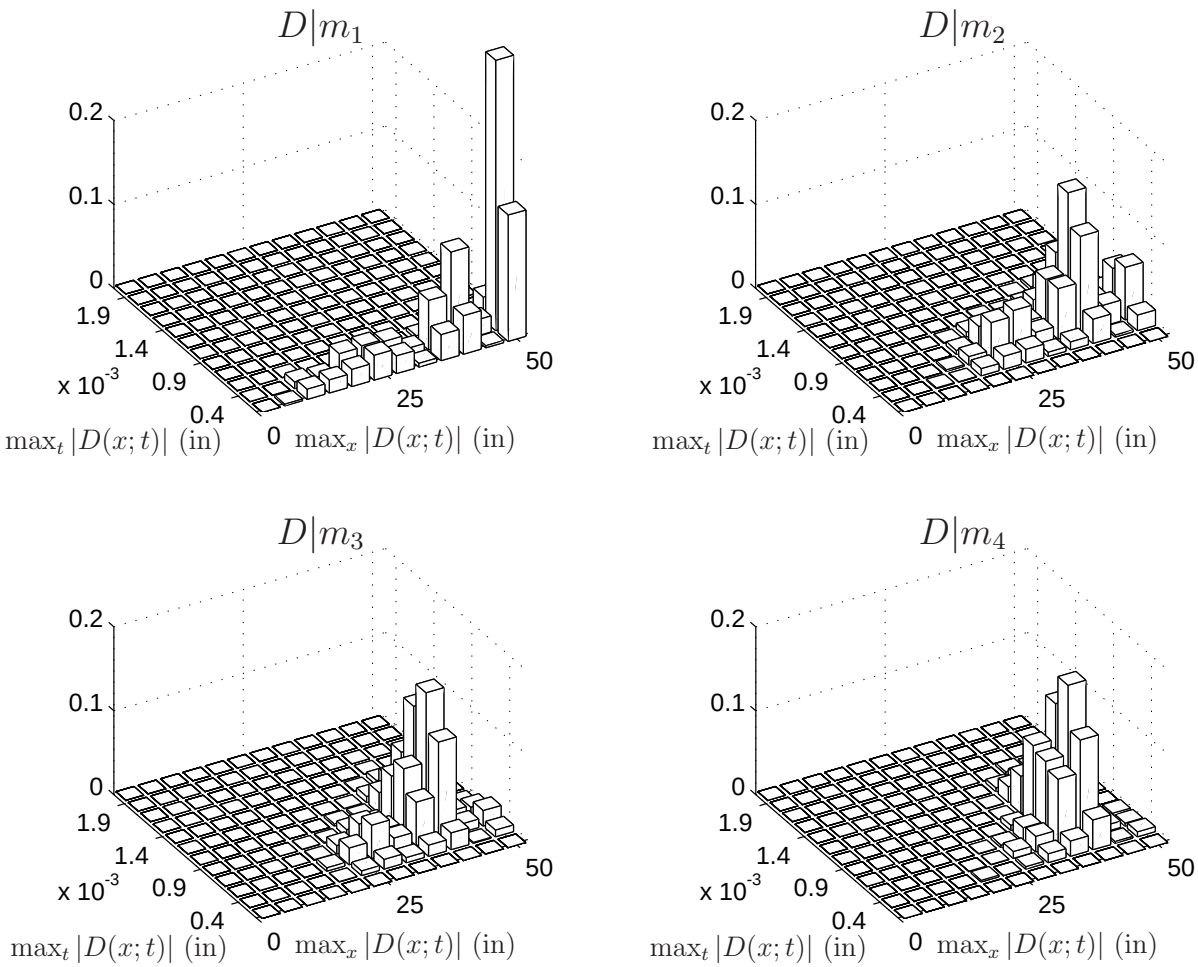

Figure 5.13: Histogram of $(\max_x |D(x;t)|, \max_t |D(x;t)|)$.

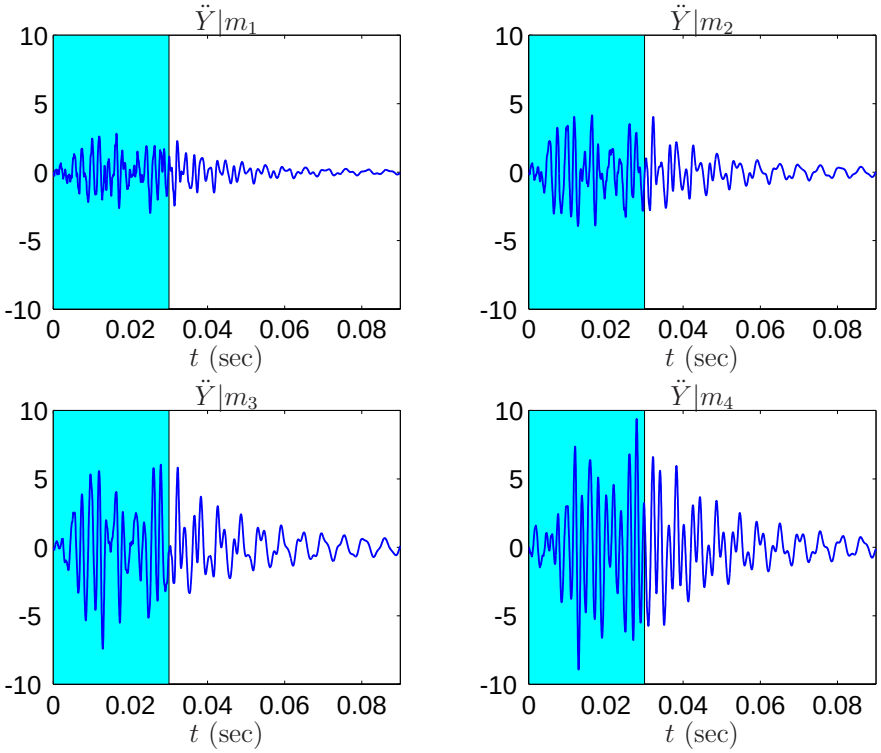

Figure 5.14: One sample of component response process, $\ddot{Y}(t)$, in units of g.

Similar patterns are observed for the acceleration response of the component, $\ddot{Y}(t)$. One sample of $\ddot{Y}(t)$ is shown in Fig. 5.14. As the degree of spatial correlation increases, so does the magnitude of the acceleration of the attached mass. Estimates of the variance of $\ddot{Y}(t)$ using results from 2,000 Monte Carlo samples are shown in Fig. 5.15 for $t \in [0, t']$. As $t \to t'$, the variance approaches a constant value for each of the four cases; this value increases with increasing spatial correlation.

Estimates of the distribution of $g(Y)$ defined by Eq. (5.16), using results from 2,000 Monte Carlo samples, are shown in Fig. 5.16. It is evident from the figure that assuming the input to be uncorrelated in space, which corresponds to the case where $\theta = 5.0$ (model m_1), is non-conservative, since the probability of exceeding a critical level of acceleration increases with decreasing θ.

For this one-dimensional model of the spacecraft, it has been demonstrated that predictions of beam displacement, $D(x;t)$, and acceleration of the mass, $\ddot{Y}(t)$, are sensitive to the degree of spatial correlation of the input, $Z(x;t)$. Specifically, as the degree of spatial correlation increases, the maximum in time of $D(x;t)$ and $\ddot{Y}(t)$ increase. In the general three-dimensional case, the input process $\mathbf{Z}(\mathbf{x};t)$ exhibits spatial correlation in two directions (axial and hoop); the sensitivity of model output demonstrated will likely increase. It is therefore crucial to select optimal models for the spatial correlation of the input.

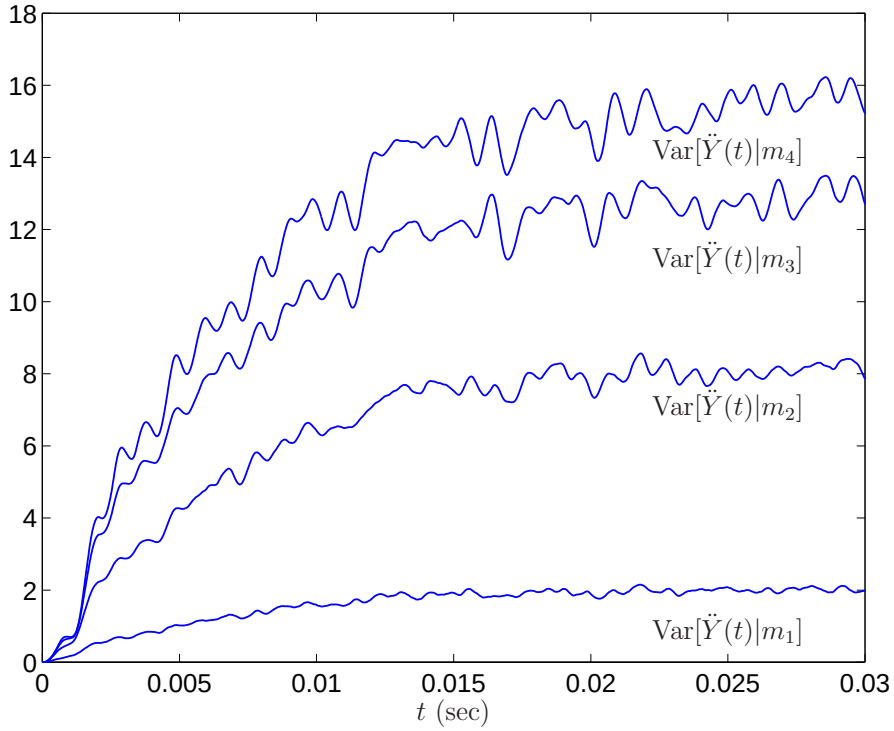

Figure 5.15: Estimates of $\mathrm{Var}[\ddot{Y}(t)]$ in units of g^2.

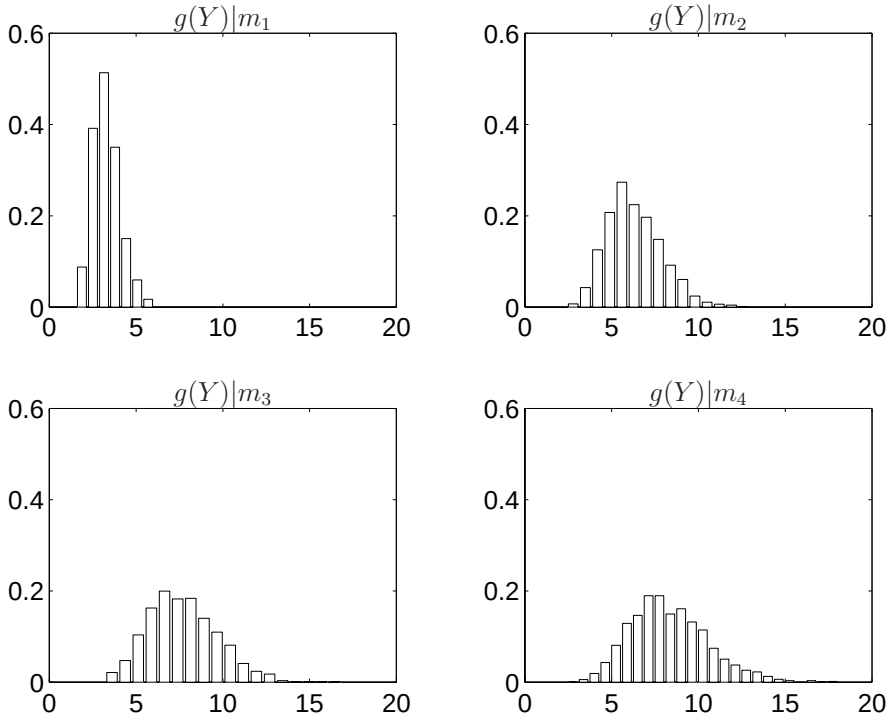

Figure 5.16: Histograms of $g(Y)$.

Table 5.3. Expected utilities of each $m_i \in \mathcal{M}'$.

d (g)	Expected utilities			
	$u(m_1)$	$u(m_2)$	$u(m_3)$	$u(m_4)$
2	7.935×10^{-4}	$\mathbf{1.323 \times 10^{-4}}$	$\mathbf{1.323 \times 10^{-4}}$	$\mathbf{1.323 \times 10^{-4}}$
5	1.114	$\mathbf{1.670 \times 10^{-1}}$	2.057×10^{-1}	2.223×10^{-1}
10	3.203×10^{-2}	2.615×10^{-2}	$\mathbf{1.108 \times 10^{-2}}$	2.287×10^{-2}

5.6.3 Optimal model

The decision-theoretic method for model selection developed in Chapter 2 is used here to select an optimal value for the spatial correlation parameter, θ. We begin with an analysis on the smaller class of candidate models, $\mathcal{M}'$ defined by Eq. (5.19), then use the results to approximate the optimal member from the larger, uncountable class, $\mathcal{M}$ defined by Eq. (5.18).

The objective is to assess the performance of the spacecraft in the re-entry environment via Eq. (5.17). Hence, the model use is prediction and the utility formulation of Section 2.3.2.3 applies. We consider the special case where the collections of candidate actions and candidate models coincide, *i.e.*, $\mathcal{A} = \mathcal{M}'$, and define the elements of the utility function as follows. The cost of taking action $m_i \in \mathcal{M}'$ is related to the computational effort required to generate samples of each model for the input process; the effort required is identical for each model so, without loss of generality, we set the cost of each model to zero, *i.e.*, $\gamma(m_i) = 0$, $i = 1, 2, 3, 4$. Let

$$h_i = P\left(g(Y) \leq d \mid m_i\right), \ i = 1, 2, 3, 4, \tag{5.20}$$

denote the prediction of performance under model m_i, where d is the specified critical value for the acceleration of the mass, $\ddot{Y}$. The consequences of prediction under model m_i when m_j is true is quantified by the penalty function, $\psi(m_i, m_j)$, given by

$$\psi(m_i, m_j) = \tilde{\psi}(h_i, h_j) = \begin{cases} \beta_1 \left[h_i - h_j\right]^2 & \text{if } h_i \leq h_j, \text{ and} \\ \beta_2 \left[h_i - h_j\right]^2 & \text{if } h_i > h_j, \end{cases} \tag{5.21}$$

which is identical to Eq. (2.73). The expected utility of each model in $\mathcal{M}'$ is then calculated via Eq. (2.35).

The expected utilities of the four models in $\mathcal{M}'$ are listed in Table 5.3 for three values of the critical acceleration, d; in each case, the optimal model, $m^\star$, is in bold. Parameters $\beta_1 = 1$ and $\beta_2 = 2$ were used for calculations, and each $m_i \in \mathcal{M}'$ is assumed equally likely, *i.e.*, $p_i = 1/4$, $i = 1, 2, 3, 4$. For $d = 5$ g and $d = 10$ g, models m_2 and m_3 are selected, corresponding to $\theta_2 = 0.1$ and $\theta_3 = 0.02$, respectively. Model m_1, corresponding to the case where the spatial correlation is nearly white noise, is

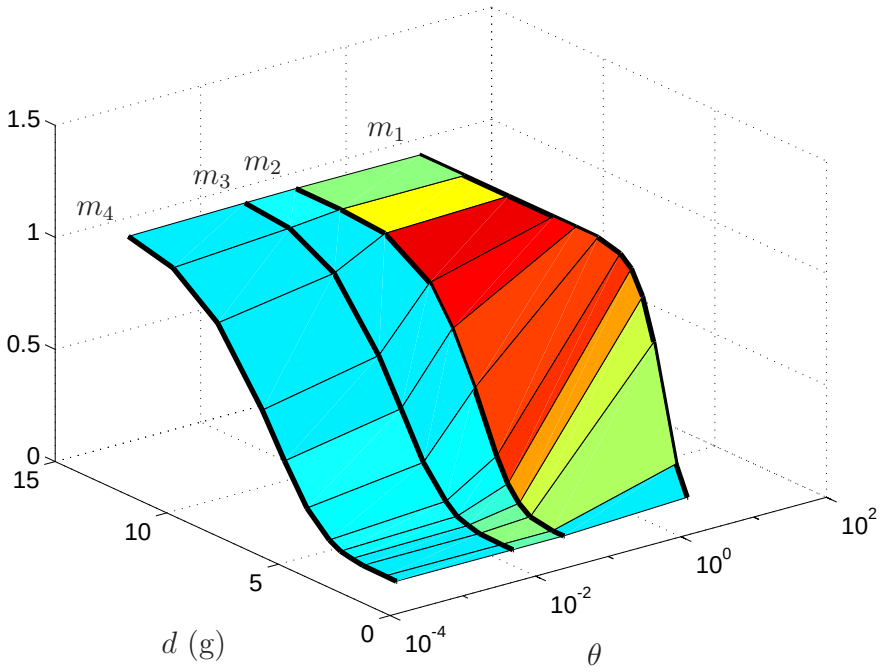

Figure 5.17: Estimates of $P(g(Y) \leq d|\theta)$ for $\theta \in \mathcal{M}$.

never selected since the assumption that the input is uncorrelated in space leads to non-conservative predictions of system performance (see Fig. 5.16). For $d = 2$ g, models m_2, m_3, and m_4 give identical estimates of the performance metric, so there is no preference of one model over another, *i.e.*, $m_2 \sim m_3 \sim m_4$ (see Section 2.3.1.3).

Estimates of $P(g(Y) \leq d|\theta)$ are plotted in Fig. 5.17 for $d \in [1, 14]$ and $\theta \in [0.0005, 5.0]$; the range for θ corresponds to $\mathcal{M}$ defined by Eq. (5.18). The estimates for the four models in $\mathcal{M}'$ are shown with a bold line; interpolation has been used to estimate $P(g(Y) \leq d|\theta)$ elsewhere. A surface defining the expected utility of each model can be defined over the values for θ and d considered; this surface is shown in Fig. 5.18, where the natural logarithm of the expected utilities are plotted to emphasize the results. The minimum of the surface for each d gives an estimate of the optimal $\theta \in \mathcal{M}$. As d increases, the optimal model corresponds to an increasing degree of spatial correlation.

5.6.4 Sensitivity of optimal model

We study the sensitivity of the optimal model to changes in the parameters of the penalty function, Eq. (5.21). To do so, define a normalized penalty function parameter $\bar{\beta} = \beta_2/\beta_1$. For $\bar{\beta} \gg 1$, non-conservative predictions of system performance are highly penalized with respect to conservative predictions; as $\bar{\beta} \to 1$, the penalty for conservative and non-conservative models is identical.

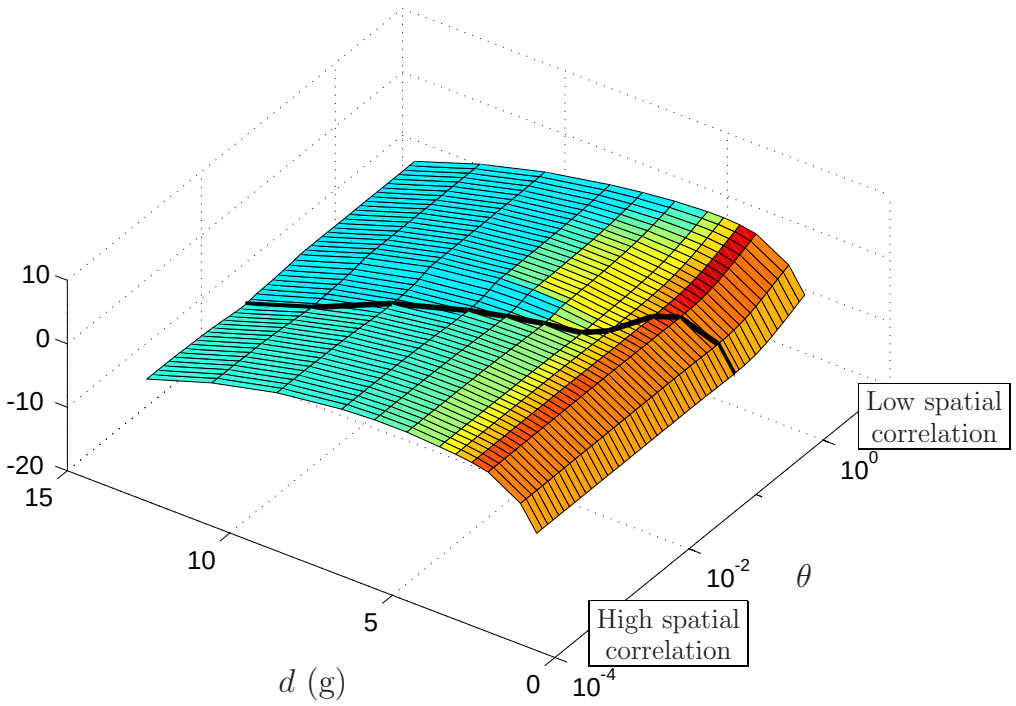

Figure 5.18: Natural log of expected utilities, $\ln[u(\theta)]$, $\theta \in \mathcal{M}$.

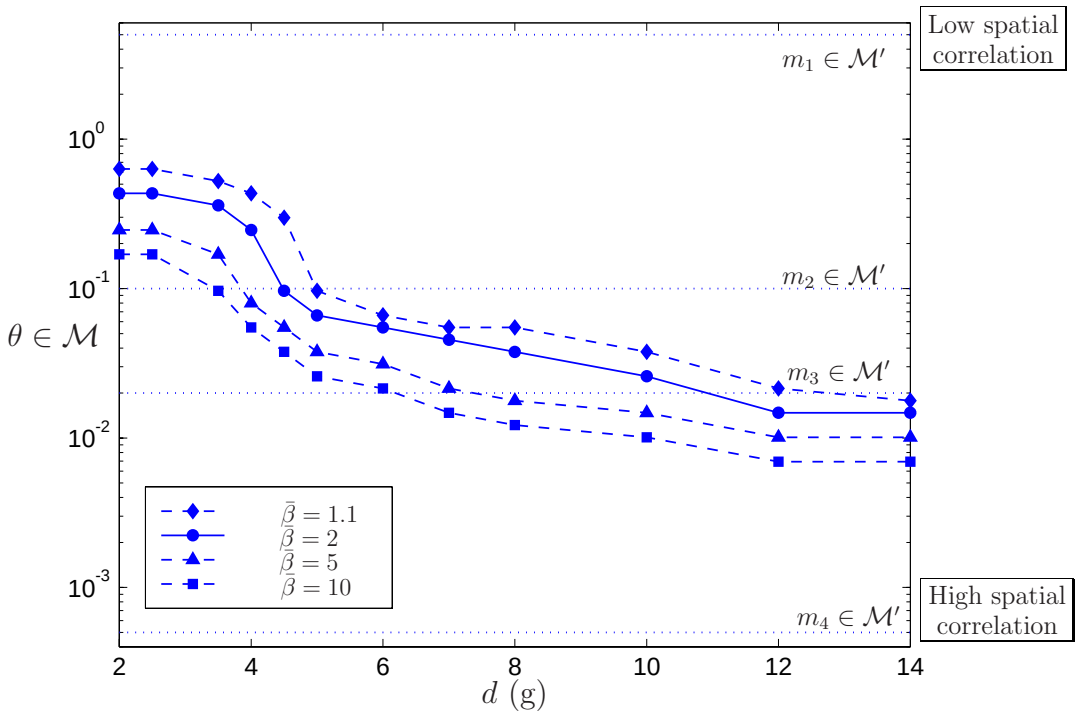

Figure 5.19: Sensitivity of optimal $\theta \in \mathcal{M}$ to changes in $\bar{\beta}$.

The optimal models from $\mathcal{M}$ are plotted in Fig. 5.19 for various values of d. Four lines are plotted, which correspond to four different values for $\bar{\beta}$. The solid line ($\bar{\beta} = 2$) corresponds to the results of the previous section. In general, changing $\bar{\beta}$ shifts the curve up or down. As $\bar{\beta}$ increases, the penalty function defined by Eq. (5.21) becomes highly asymmetrical, and models with more spatial correlation (smaller θ) are favored. The dotted lines correspond to the four models from $\mathcal{M}'$. In all cases, the optimal spatial correlation parameter satisfies $0.005 \leq \theta \leq 0.5$, meaning that the limiting cases of perfect or zero spatial correlation are unfavorable, even for large changes in $\bar{\beta}$.

5.7 Case #2: model selection for form of spatial correlation function

In this section, we again consider the input pressure field defined by Eq. (5.10), where $Q = G$ is a stationary Gaussian process with zero mean, unit variance, and temporal correlation function given by Eq. (5.13). Instead of the fixed functional form for the spatial correlation function studied in the previous section, here we consider models for $\phi(\eta;\theta)$ with different functional forms. The spatial correlation parameter for each of these models, θ, is also unknown.

5.7.1 Candidate models

The functional form of ϕ, as well as the spatial correlation parameter, θ, are unknown. The class of candidate models is thus much larger than for the case of ϕ with known, fixed functional form given by Eq. (5.18). In general, the class of candidate models for this problem is

$$\mathcal{M} = \{\mathcal{M}_i\},\ i \in I, \tag{5.22}$$

where each $\mathcal{M}_i = \{\theta_i : \phi_i(\eta;\theta_i)\}$, and I is an uncountable index set. With $I = \{1\}$ and $\phi_1 = \exp(-\theta_1|\eta|)$, Eq. (5.22) reduces to the class of models considered in the previous section, Eq. (5.18). In this section, we consider the special case $I = \{1,2,3\}$, so that $\mathcal{M} = \{\mathcal{M}_1, \mathcal{M}_2, \mathcal{M}_3\}$, where

$$\begin{aligned}
\mathcal{M}_1 &= \{\theta_1 \colon \phi_1(\eta;\theta_1) = \exp(-\theta_1 \mid \eta \mid)\}, \\
\mathcal{M}_2 &= \{\theta_2 \colon \phi_2(\eta;\theta_2) = \exp\left[-(\theta_2\eta)^2\right]\}, \\
\mathcal{M}_3 &= \{\theta_3 \colon \phi_3(\eta;\theta_3) = \exp(-\theta_3 \mid \eta \mid)\cos(2\theta_3\eta)\}.
\end{aligned} \tag{5.23}$$

The three functional forms of Eq. (5.23) are plotted in Fig. 5.20 for $\theta_1 = \theta_2 = \theta_3 = 0.1$. Note that ϕ_1 was used in the previous section; functional forms ϕ_2 and ϕ_3 have been used extensively to represent, among other phenomena, the spatial correlation of seismic ground motions [99].

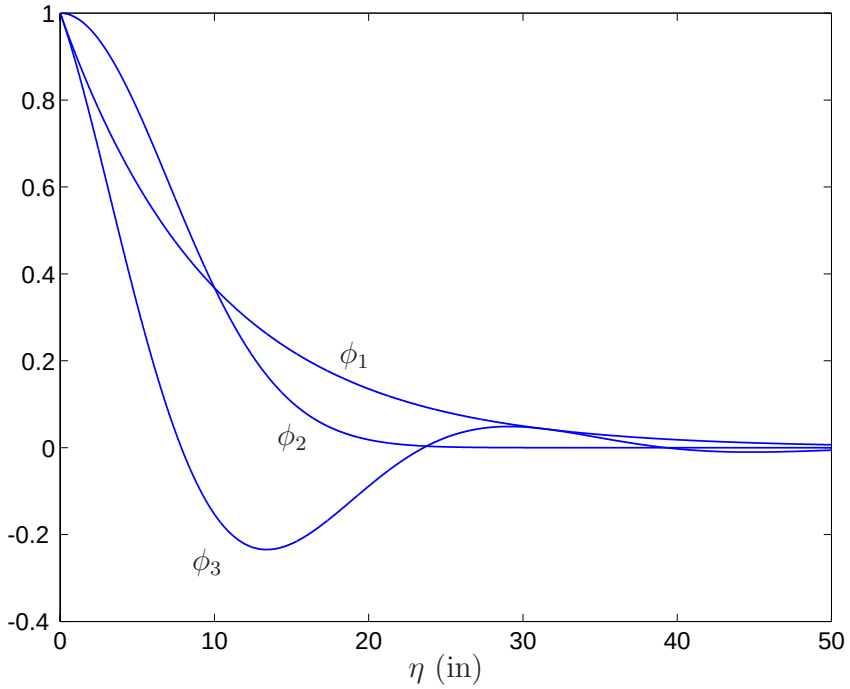

Figure 5.20: Candidate functional forms for spatial correlation, ϕ.

5.7.2 Optimal model

We next summarize the procedure to select the optimal model in $\mathcal{M}$ defined by Eq. (5.22). First, we use the techniques described in the previous section to estimate the optimal parameter, θ_1, of ϕ_1; this is simply the model selection problem applied to $\mathcal{M}_1$ defined by Eq. (5.23) and is identical to the analysis of Section 5.6. We then repeat this process to estimate optimal parameters θ_2 and θ_3 of functional forms ϕ_2 and ϕ_3, respectively. At this point, the problem simplifies to choosing the optimal member from

$$\mathcal{M} = \{m_i\} = \{\phi_i(\eta; \theta_i)\}, i = 1, 2, 3, \tag{5.24}$$

where each θ_i is known and fixed. Methods of model selection can be applied one final time to choose the optimal $m_i \in \mathcal{M}$. A schematic of this procedure is shown in Fig. 5.21.

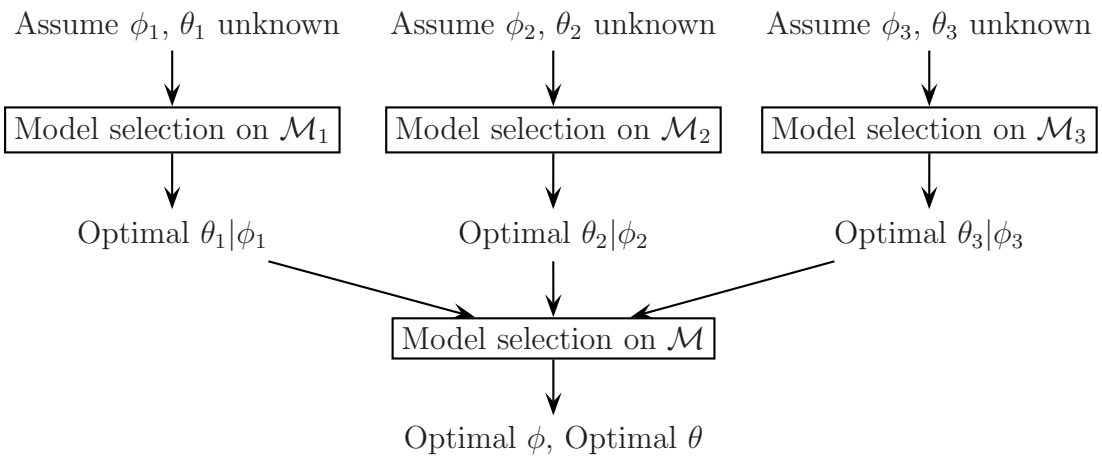

Figure 5.21: Overview of model selection procedure for Section 5.7.

5.8 Case #3: model selection for marginal PDF of input

Recall Figs. 5.1 and 5.2, which showed measured accelerations during the re-entry of the Space Shuttle Orbiter, and wind-tunnel pressure measurements on an aircraft fuselage, respectively. In both plots, non-Gaussian behavior is clearly evident. In this section, we consider non-Gaussian models for the input, Z; methods of model selection are used to select the optimal member from a class of candidate models for Z.

Two assumptions are made in this section: (1) ϕ is perfectly known and equal to one, and (2) the dynamics of the beam and oscillator can be decoupled. Assumption (1) implies that Eq. (5.10) reduces to

$$Z(x;t) = \sigma_Z(x)Q(t), \ t \in [0, t'], \tag{5.25}$$

where the dependence of Q on x is removed because it exhibits perfect spatial correlation. As a result, the second-moment properties of Z are now completely defined. All candidate models considered in this section are equivalent in the second-moment sense, but have different marginal distributions. Because of assumption (2), the method for analysis is given by Eq. (5.9). This allows us to study the effects of the distribution of Z on $\ddot{Y}$ for various values of b, the location of the oscillator (see Fig. 5.5), and ω_c, the natural frequency of the oscillator.

Table 5.4 lists the values of the oscillator parameters used in the analyses that follow. The damping ratio of the oscillator, $\zeta_c = 0.02$, as well as the parameters of the beam from Table 5.2, remain fixed. Note that one of the resonant frequencies of the oscillator considered, $\omega_c = 2{,}000$ Hz, coincides with a resonant frequency, ν_7, of the beam from Fig. 5.7.

In Section 5.8.1, we define a collection of candidate models for Z. We illustrate the

Table 5.4. Oscillator parameters.

Parameter	Values	Units
b	12.5, 22.5, 37.5	in
ω_c	2000, 2250, 5000	Hz

sensitivity of the acceleration of the mass to different models for Z in Section 5.8.2. The optimal model for Z, denoted by $m^\star$, as well as the sensitivity of $m^\star$ to changes in the utility function are discussed in Sections 5.8.3 and 5.8.4, respectively.

5.8.1 Candidate models

Let

$$\mathcal{M}' = \{m_i\} = \{Z_i(x;t)\},\ i = 1, 2, 3, 4, \tag{5.26}$$

be the class of candidate models to be considered for the applied pressure field, where

$$Z_i(x;t) = \sigma_Z(x) Q_i(t). \tag{5.27}$$

Define $Q_1 = G$, where G is a zero mean, unit variance, stationary Gaussian process with correlation function given by Eq. (5.13). This model is identical to the model considered in Section 5.6 for $\theta = 0$. Model m_1 is Gaussian.

Let Q_2 be a translation process with target marginal CDF, F, *i.e.*,

$$Q_2(t) = F^{-1} \circ \Phi[G(t)]. \tag{5.28}$$

Here, F is taken to be the student-t distribution with $(r-1)$ degrees-of-freedom [5], *i.e.*,

$$F(z;r) = \frac{\Gamma\left(\frac{r+1}{2}\right)}{\sqrt{\pi r}\ \Gamma\left(\frac{r}{2}\right)} \int_{-\infty}^{\sqrt{\frac{r}{r-2}}z} \left[1 + \frac{t^2}{r}\right]^{-1/2(r+1)} \mathrm{d}t, \tag{5.29}$$

where $\Gamma(\,\cdot\,)$ denotes the Gamma function. A random variable with the distribution given by Eq. (5.29) has zero mean and unit variance; the process $Z_2(x;t) = \sigma_Z(x) Q_2(t)$ is therefore equivalent in the second-moment sense to $Z_1(x;t)$ for any r. $Z_2(x;t)$ can be assigned any coefficient of kurtosis greater than 3 without altering the second-moment properties via [59]

$$\mathrm{Kur}[Z_2] = \frac{3(r-2)}{r-4},\ r > 4. \tag{5.30}$$

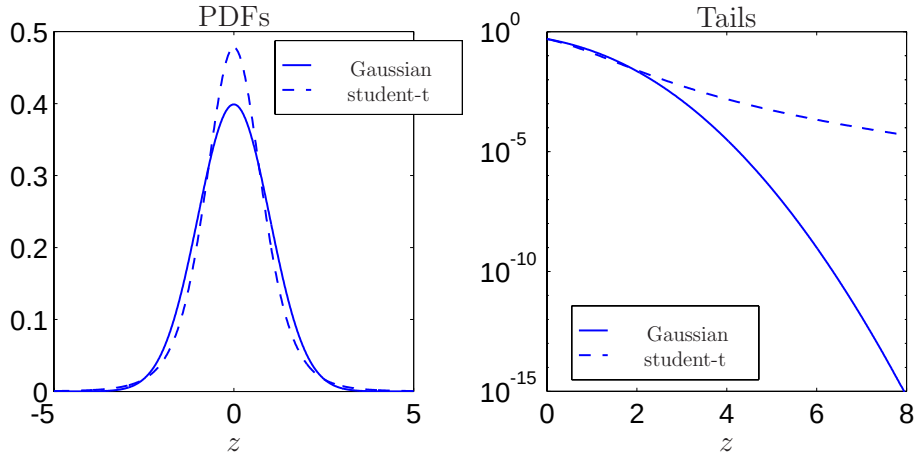

Figure 5.22: PDFs (left) and tails (right) of Gaussian and student-t distributions.

The PDF of the student-t distribution with a coefficient of kurtosis of 7 ($r = 11/2$), is plotted along with the Gaussian PDF in the left-hand side of Fig. 5.22. This distribution has heavier tails that that of the Gaussian distribution, as illustrated further by the right-hand side of the figure; the Gaussian distribution has a coefficient of kurtosis of 3.

The third and fourth models considered for the applied pressure field are $Q_3(t) = V(t;\lambda_3)$ and $Q_4(t) = V(t;\lambda_4)$, respectively, where

$$V(t;\lambda) = \begin{cases} 0 & N(t) = 0 \\ \sum_{k=1}^{N(t;\lambda)} W_k f(t - T_k) & N(t) > 0. \end{cases} \tag{5.31}$$

The process $V(t)$ is a type of **filtered Poisson process**, a process characterized by pulses of deterministic shape and random magnitude, occurring at (random) Poisson times. $V(t)$ depends on $\{N(t;\lambda), t \geq 0\}$, a homogeneous Poisson counting process of intensity $\lambda > 0$, iid random variables $\{W_k, k \geq 1\}$, with mean zero and variance σ^2, the random times, $\{T_k, k \geq 1\}$, at which the Poisson events occur, and the deterministic shape function, $f(t)$, $t \geq 0$ [48]. The process $N(t)$ has the property that $\mathrm{E}[N(t)] = \lambda t$ so that $\lambda = \mathrm{E}[N(t)]/t$ represents the average number of pulses per unit time [83]. With proper choice of $f(t)$, σ^2, and λ, models $Z_3(x;t)$ and $Z_4(x;t)$ are equivalent to $Z_1(x;t)$ and $Z_2(x;t)$ in the second-moment sense (see Appendix C). For small λ, the pulses have large magnitudes and occur infrequently. As λ increases, the pulses occur more frequently, but with smaller magnitude; as $\lambda \to \infty$, the marginal distribution of $V(t)$ becomes Gaussian [48].

One sample of the Gaussian process, $Q_1(t) = G(t)$, the translation process, $Q_2(t) = F^{-1} \circ \Phi[G(t)]$, and the two filtered Poisson processes, $Q_3(t) = V(t;\lambda_3)$ and $Q_4(t) = V(t;\lambda_4)$ with $\lambda_3 = 10{,}000$ and $\lambda_4 = 100{,}000$, respectively, are plotted in Fig. 5.23. Samples from the four models look quite different, but each is consistent

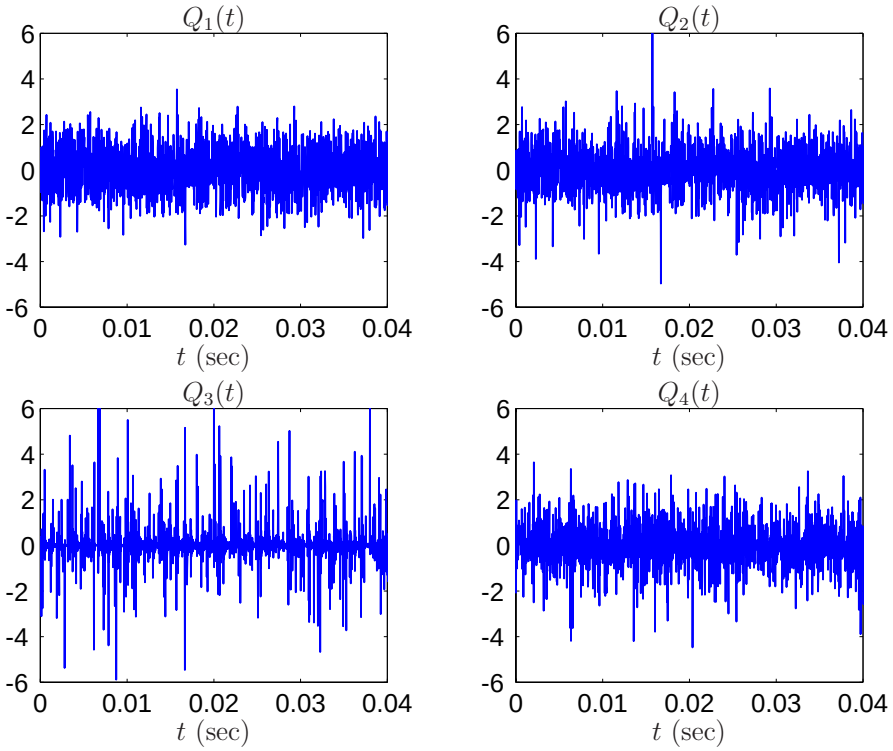

Figure 5.23: One sample of the four models for Q.

with all available information. Estimates of the marginal PDFs of Q_3 and Q_4 using 500 samples are shown in Fig. 5.24. Also shown is the marginal PDF of Q_1.

5.8.2 Sensitivity of model output

One sample of the acceleration of the mass, $\ddot{Y}(t)$, is shown in Fig. 5.25; the four plots correspond to the four models of the applied pressure field for the case of $b = 12.5$ in and $\omega_c = 5{,}000$ Hz. The shaded area indicates the time interval when the load is applied, $t \in [0, t' = 0.05 \text{ sec}]$; the unshaded area corresponds to free vibration, $t \in [t', 3t']$.

$\ddot{Y}$, is the output from a linear filter. Because of this, and the fact that all inputs considered are equivalent in the second-moment sense, the outputs will have identical second-moment properties [48]. This is further illustrated by Fig. 5.26, which shows estimates of the variance of $\ddot{Y}(t)$, $t \in [0, t']$, using 1,000 samples, for the case of $b = 12.5$ in, $\omega_c = 5{,}000$ Hz. The variance of $\ddot{Y}$ is computed assuming each of the four models in $\mathcal{M}'$ are the input.

As previously discussed, the probability laws for $m_i \in \mathcal{M}'$ are different; the resulting probability laws for $\ddot{Y}(t)$, and hence the performance metric defined by Eq. (5.17)

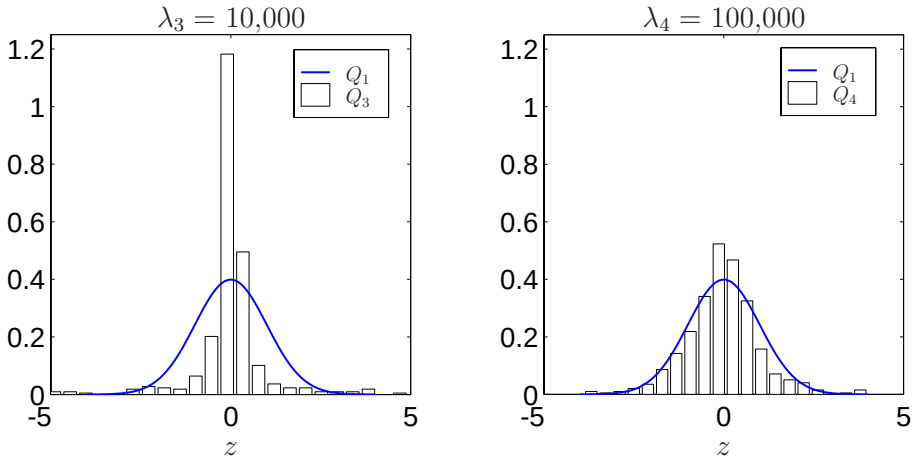

Figure 5.24: Marginal PDFs of Q_3 (left) and Q_4 (right).

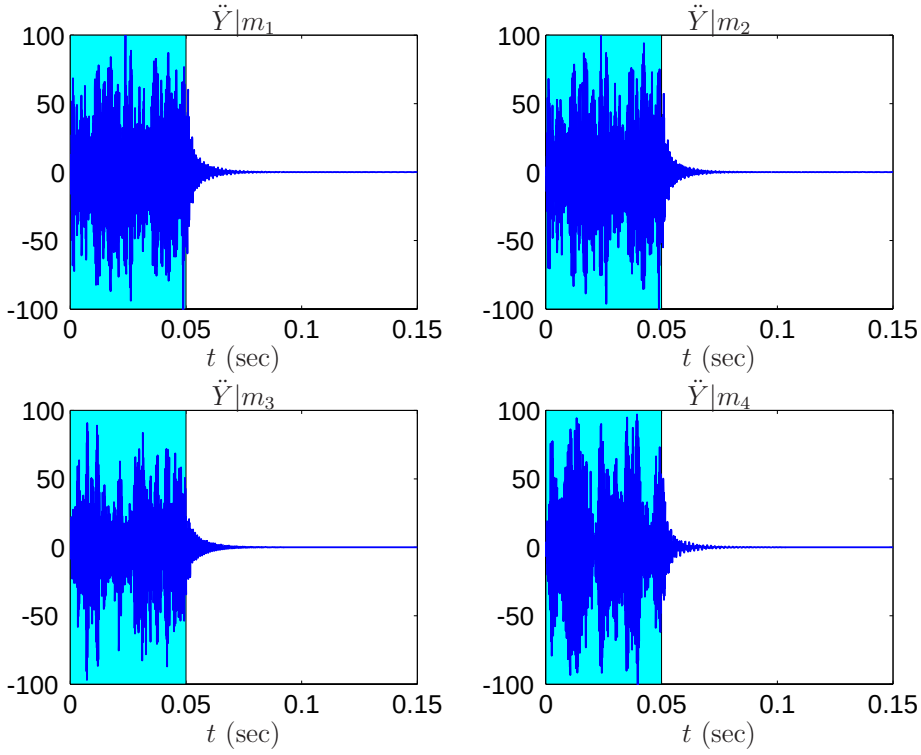

Figure 5.25: One sample of $\ddot{Y}(t)$ under each $m_i \in \mathcal{M}'$.

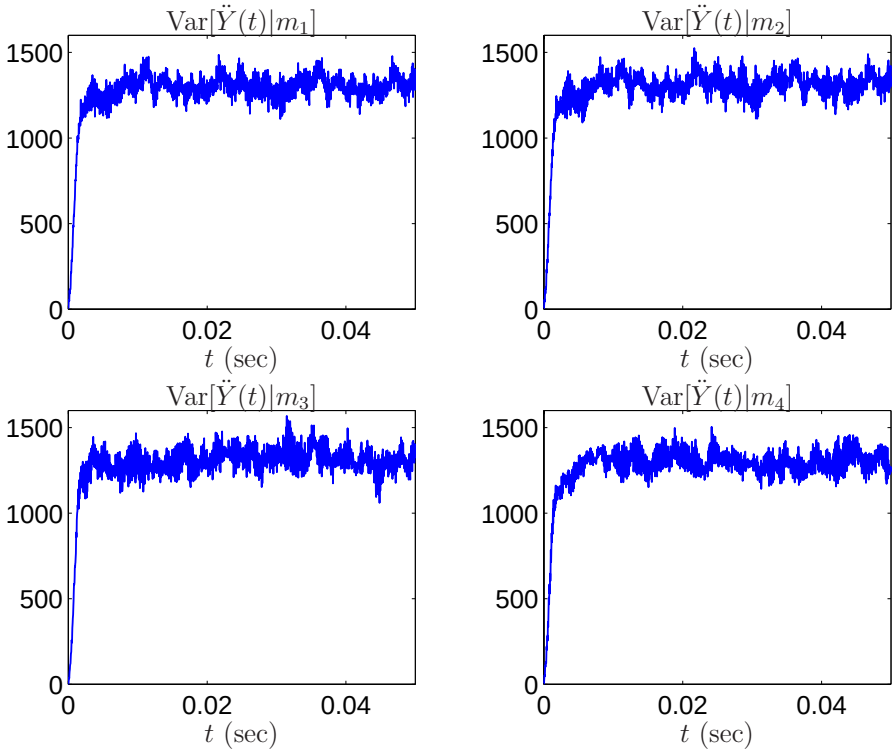

Figure 5.26: Estimates of $\mathrm{Var}[\ddot{Y}(t)]$, $t \in [0, t']$.

will therefore also differ. To illustrate, we define

$$h_i(\hat{\sigma}d) = P\left(g(Y) \leq \hat{\sigma}d \mid m_i\right), \ i = 1, 2, 3, 4. \tag{5.32}$$

The quantity $\hat{\sigma}$ is the estimated steady-state standard deviation of $\ddot{Y}$ taken from Fig. 5.26, and is used to normalize the tail estimates; $g(Y)$ is defined by Eq. (5.16). Estimates of Eq. (5.32) are illustrated by Fig. 5.27 for two cases: (1) $b = 37.5$ in, $\omega_c = 2{,}000$ Hz (left), and (2) $b = 12.5$ in, $\omega_c = 5{,}000$ Hz (right). As shown by the plot on the right, models m_2, m_3, and m_4, in general, provide conservative estimates of Eq. (5.32) with respect to the Gaussian model, m_1. Model m_3 is the most conservative, due to the fact that Z_3 exhibits the heaviest tail. For $\omega_c = 2{,}000$ Hz, a resonant frequency of the beam, the tail estimates are nearly identical, as shown in the left-hand side of Fig. 5.27. The response of the oscillator is insensitive to the tails of $Z(x;t)$ when ω_c coincides with a resonant frequency of the beam.

The magnitude of the difference between the tails of the distributions shown in Fig. 5.27 is sensitive to changes in parameters b and ω_c. To illustrate, we define

$$\rho(\hat{\sigma}d) = \max_i \left|h_1(\hat{\sigma}d) - h_i(\hat{\sigma}d)\right|, \ i = 2, 3, 4. \tag{5.33}$$

Equation (5.33) is plotted on the left-hand side of Fig. 5.28 for the case of $\hat{\sigma}d = 4$ and different values for b and ω_c. The color of the bars denotes which model maximizes ρ; in seven of the nine cases considered, it is m_3 (white), while it is m_2 in the remaining two cases (gray). In general, for increasing b, ρ may increase or decrease. As the

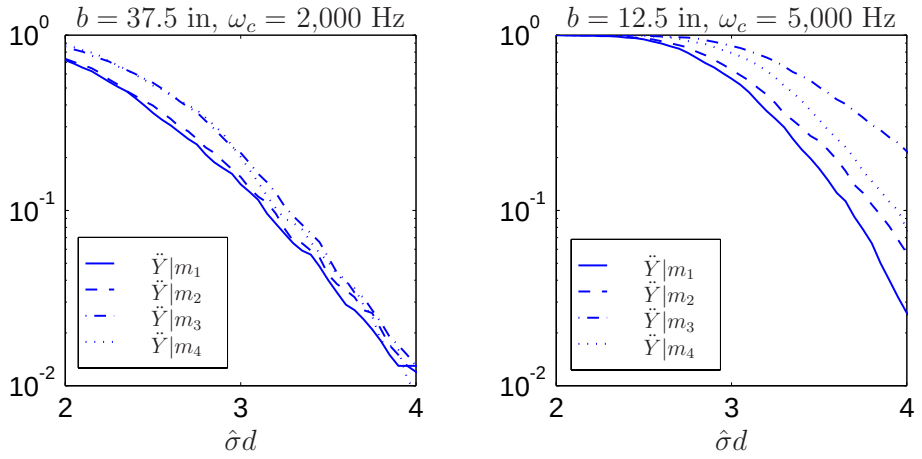

Figure 5.27: Estimates of $1 - h_i(\hat{\sigma}d)$, $i = 1, 2, 3, 4$.

frequency of the oscillator is increased, ρ increases; for $\omega_c = 2{,}000$ Hz, a resonant frequency of the beam, ρ is near zero. Hence, the sensitivity to the tails of Z can be minimized by placing ω_c near a resonant frequency of the beam. Of course, this has the undesirable effect of increasing the variance of the response, as depicted by the right-hand side of Fig. 5.28.

It has been demonstrated that predictions of component acceleration, $\ddot{Y}$, are sensitive to the tails of the marginal distribution of the input, Z, when the oscillator natural frequency, ω_c, does not coincide with a resonant frequency of the beam. As the tail of the marginal PDF of the input model becomes heavier, the maximum in time of $\ddot{Y}(t)$ increases. A Gaussian model for Z can therefore give non-conservative predictions of system performance.

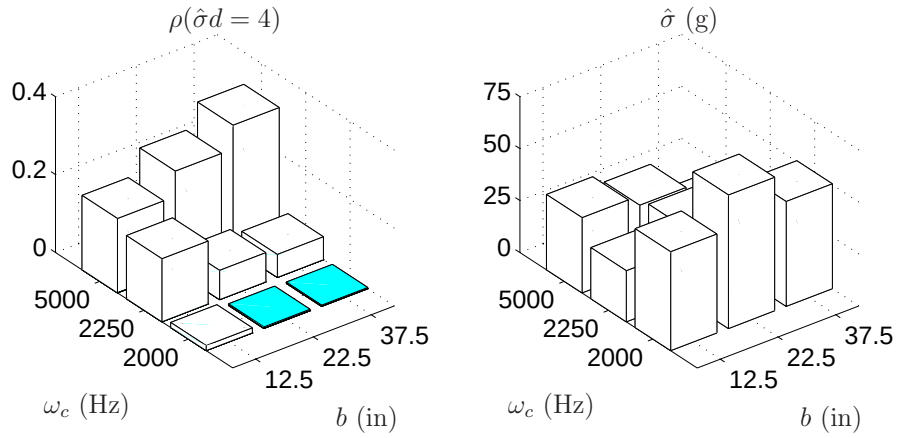

Figure 5.28: Estimates of ρ and $\hat{\sigma}$ for different values for b and ω_c.

Table 5.5. Average CPU times and costs.

Model	CPU time (sec)	Cost, $\gamma(m_i)$
m_1	0.6558	γ_0
m_2	1.025	$1.563\gamma_0$
m_3	0.6886	$1.050\gamma_0$
m_4	6.114	$9.323\gamma_0$

5.8.3 Optimal model

In this section, we use the decision-theoretic method for model selection developed in Chapter 2 to select the optimal marginal distribution of the input pressure field from the class of candidate models defined by Eq. (5.26).

The penalty function used in the analysis is unchanged from the previous section, *i.e.*, Eq. (5.21). However, the computational effort associated with each model in $\mathcal{M}'$ is no longer identical; the costs, $\gamma(m_i)$, $i = 1, 2, 3, 4$, must reflect this. Table 5.5 lists the average CPU times required to calculate one sample of each of the four models in $\mathcal{M}'$. Samples of the Gaussian process, m_1, can be calculated most rapidly; the filtered Poisson process with $\lambda_3 = 10{,}000$, m_3, requires slightly longer. Samples of the filtered Poisson process with $\lambda_4 = 100{,}000$, m_4, take approximately ten times longer to generate than either m_1 or m_3. The corresponding costs, scaled by deterministic parameter $\gamma_0 > 0$, are listed in Table 5.5. In each case, $\gamma(m_i)/\gamma_0$ is the average CPU time of m_i, normalized by the CPU time of m_1 (the fastest).

Table 5.6 lists the expected utilities of each $m_i \in \mathcal{M}'$ using Eq. (2.35). Results for the different values of b, the location of the oscillator, and ω_c, the resonant frequency of the oscillator, are shown for $\hat{\sigma}d = 4$. Parameters $\beta_1 = 1$, $\beta_2 = 2$, and $\gamma_0 = 0.05$ were used for calculations. The translation process, m_2, is selected most often; the cost of m_2 is not significantly greater than that of m_1, and it, in general, provides slightly conservative estimates of $P(g(Y) \leq d)$. The filtered Poisson process with $\lambda = 10{,}000$, m_3, is selected only twice; it provides overly conservative estimates of system performance in many cases. The Gaussian model, m_1, is selected only for the case when $\omega_c = 2{,}000$ Hz, *i.e.*, when the oscillator frequency coincides with a resonant frequency of the beam. This is because the tail estimates nearly coincide here (see Fig. 5.27), and the cost term dominates the utility. Model m_4 provides slightly conservative estimates of system performance, but is never selected due to its large computational cost (see Table 5.5).

Table 5.6. Expected utilities of each $m_i \in \mathcal{M}$.

b (in)	ω_c (Hz)	$u(m_1)$	$u(m_2)$	$u(m_3)$	$u(m_4)$
12.5	2,000	$\mathbf{5.109 \times 10^{-3}}$	7.924×10^{-3}	5.344×10^{-3}	4.666×10^{-2}
12.5	2,250	1.854×10^{-2}	$\mathbf{1.800 \times 10^{-2}}$	2.182×10^{-2}	5.667×10^{-2}
12.5	5,000	2.507×10^{-2}	$\mathbf{2.108 \times 10^{-2}}$	2.560×10^{-2}	5.706×10^{-2}
22.5	2,000	$\mathbf{5.013 \times 10^{-3}}$	7.830×10^{-3}	5.265×10^{-3}	4.667×10^{-2}
22.5	2,250	8.294×10^{-3}	1.006×10^{-2}	$\mathbf{8.209 \times 10^{-3}}$	4.796×10^{-2}
22.5	5,000	4.013×10^{-2}	$\mathbf{3.061 \times 10^{-2}}$	4.317×10^{-2}	6.811×10^{-2}
37.5	2,000	$\mathbf{5.003 \times 10^{-3}}$	7.819×10^{-3}	5.254×10^{-3}	4.664×10^{-2}
37.5	2,250	7.085×10^{-3}	9.276×10^{-3}	$\mathbf{7.060 \times 10^{-3}}$	4.742×10^{-2}
37.5	5,000	5.810×10^{-2}	$\mathbf{3.972 \times 10^{-2}}$	6.699×10^{-2}	9.117×10^{-2}

5.8.4 Sensitivity of optimal model

We consider the sensitivity of the optimal model to changes in parameter γ_0, the weight on the cost. For $\gamma_0 = 0$, the computational costs of the models are ignored in the model selection process; as $\gamma_0 \to \infty$, the cost dominates the model selection process. Figure 5.29 shows the optimal model, denoted by $m^\star$, for each value of b and ω_c considered, for changing γ_0. The case where $\gamma_0 = 0.005$ corresponds to the results of Table 5.6. For $\gamma_0 = 0$, computational cost is ignored, and model m_4 is selected most often. As γ_0 increases, the Gaussian model, m_1, is optimal for any b and ω_c.

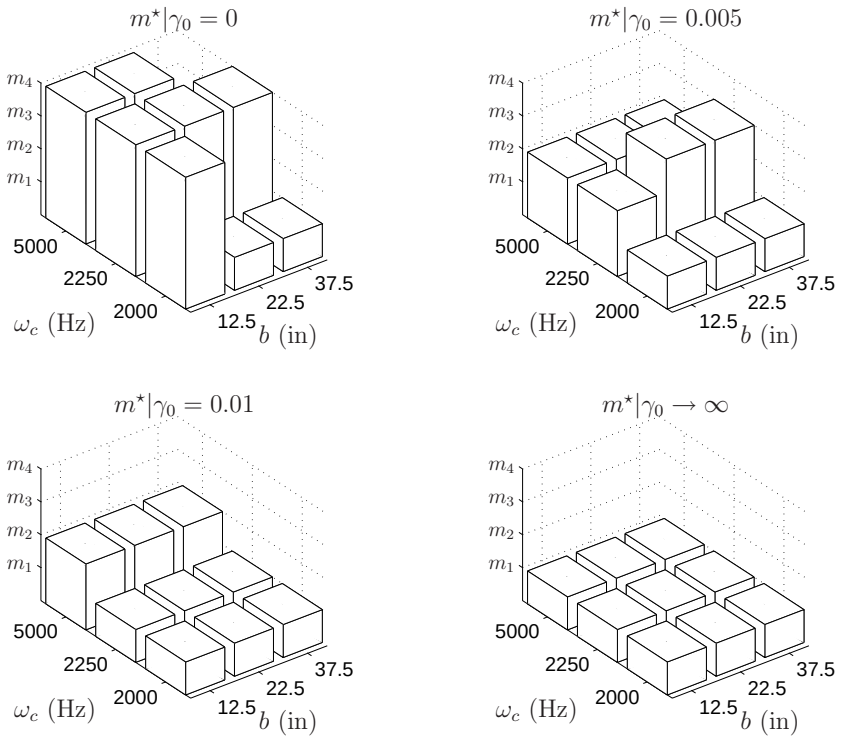

Figure 5.29: Sensitivity of $m^\star$ to changes in cost parameter, γ_0.

This page intentionally left blank.

Chapter 6

Design of sensor networks for vehicle tracking and identification

The tracking and identification of real-world moving objects for the purpose of surveillance is a widespread application [19, 20, 96]. Examples include air traffic control, fleet tracking, habitat monitoring, mobile telephony, and military battle planning [9]. There are numerous types of sensing devices that can be used for surveillance; examples include those that measure acoustic, seismic, and/or electromagnetic disturbances [65]. For surveillance, a network of one or more of these sensing devices, herein called a **sensor network**, is used to survey the environment. Observations by the sensor network are then used to make decisions regarding the identity of objects, as well as the trajectories these objects may take.

In a typical application, the sensor network is designed to: (1) **detect** any real-world objects, (2) **classify** which objects are of interest, and (3) **monitor** the locations of these objects. However, several factors exist that can complicate the design problem; three are discussed here. First, each sensor in the network reports measurements from diverse sources including, but not limited to, objects of interest, objects of little or no interest, background noise sources such as clutter, or measurement error sources such as thermal noise [12]; these ancillary sources may confuse the sensor. Second, each sensor in the network has a finite range or field of view, making it impossible to monitor all objects at all times. Third, the design of the network is subject to cost constraints, which limits both the number of sensors that can be included in the design, as well as the accuracy of each sensor. Due to these complicating factors, objects may be misclassified, *e.g.*, an object may be classified to be of interest when it is not, or even go undetected; the ability of the sensor network to adequately monitor objects may also be adversely affected by these factors.

Available information on the complicating factors listed above is typically limited, and we are forced to make assumptions about what is unknown in order to design the sensor network. The implications of this can be severe. For example, consider the

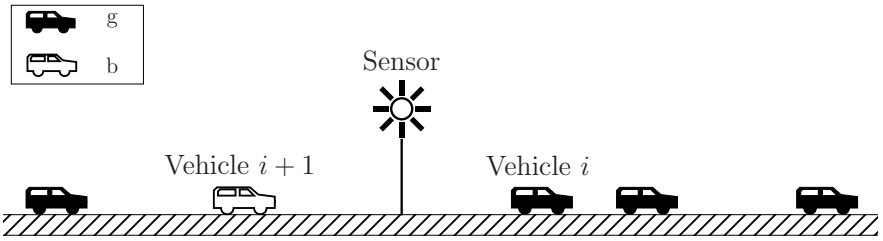

Figure 6.1: One-dimensional roadway for vehicle detection and classification.

case where there is sufficient information to model the sensor measurement error as a sequence of zero-mean iid random variables; the available information on the variance of this sequence, however, is incomplete. The design of the sensor network, and hence the ability to detect, classify, and monitor any objects of interest, may be sensitive to any assumption on the variance of the measurement error. If a poor assumption is made, the conclusions made by the network about the identity and trajectory of objects may be highly inaccurate. For this reason, the application discussed in this chapter is also classified as a high risk system.

We use the decision-theoretic method for model selection developed in Chapter 2 to choose the optimal design of a sensor network for vehicle surveillance. The three elements of the design, namely the detection, classification, and monitoring of objects, are treated independently; the extension to the general case where all elements are treated simultaneously is straightforward. In Section 6.1, we choose the optimal level of sensitivity for a single sensor to detect and classify vehicles. The design of a network with two or more sensors to monitor the movement of vehicles in a given region is discussed in Section 6.2; the objective is to select the optimal number of sensors, their locations, and the range of each.

6.1 Vehicle detection and classification

Consider the one-dimensional roadway illustrated by Fig. 6.1, where vehicles move from left to right, passing a single sensor. It is assumed that: (1) there is a single sensor located in a fixed position on the roadway, (2) all vehicles travel with constant, known velocity in the same direction, and (3) each vehicle can be characterized as one of two possible types: **g** (good) or **b** (bad). The objective is to design the sensor to activate if/when, and only if/when, a type **b** vehicle passes.

In reality, the sensor cannot function perfectly, *e.g.*, the sensor will be activated by some type **g** vehicles, and/or some type **b** vehicles will pass without activating the sensor. The presence of background noise and/or measurement error will increase the frequency of occurrence these imperfections. In general, four events can occur; these are described by the array in Table 6.1. The $(1, 1)$ element of the array is a success,

Table 6.1. Possible scenarios for vehicle detection and classification.

Sensor is	**Type b vehicle**	
	does pass	does not pass
activated	detection	false alarm
not activated	miss	-

we define it as

$$\text{detection} = \{ \text{ sensor is activated} \mid \text{type } \mathbf{b} \text{ vehicle does pass} \}. \tag{6.1}$$

Two types of failure are possible; they are defined as

$$\begin{aligned} \text{false alarm} &= \{ \text{ sensor is activated} \mid \text{type } \mathbf{b} \text{ vehicle does not pass} \}, \text{ and} \\ \text{miss} &= \{ \text{ sensor is not activated} \mid \text{type } \mathbf{b} \text{ does pass} \}, \end{aligned} \tag{6.2}$$

which correspond to the the $(1, 2)$ and $(2, 1)$ elements of the array in Table 6.1, respectively. The objective of the sensor design problem can be re-stated as follows: design the sensor such that the events defined by Eqs. (6.1) and (6.2) satisfy some performance requirements.

Mathematical models for both the vehicle traffic and sensor are developed in Section 6.1.1, where the model of the latter may include effects of measurement error. Metrics that define sensor performance are developed in Section 6.1.2. The design and analysis of the sensor, using methods of model selection, is then considered with and without the presence of measurement error; the results are discussed in Sections 6.1.3 and 6.1.4, respectively. The design and analysis of the sensor in the presence of background noise is discussed in Section 6.1.5.

6.1.1 Model development

In this section, we formulate the mathematical model for the vehicle traffic and for the sensor. This includes models for: (1) the vehicle arrival times and attributes, and (2) the sensor output and classification rule.

6.1.1.1 Vehicle model

Let $N(t)$ be a stochastic process that governs the total number of vehicles that pass the sensor in the time interval $[0, t]$, and let $T_1, T_2, \ldots, T_{N(t)}$ be the random times at

which the vehicles arrive at the sensor. Here, we assume $N(t)$ to be a homogeneous Poisson counting process with parameter $\lambda > 0$ [50], *i.e.*,

$$N(t) = \sum_{n \geq 1} 1(t \geq T_n),\ t \geq 0, \tag{6.3}$$

where

$$P(N(t) = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t},\ n = 0, 1, 2, \ldots, \tag{6.4}$$

is the probability law of the random variable $N(t)$. Note that $\mathrm{E}[N(t)] = \lambda t$, so that the average number of vehicles passing the sensor in $[0, t]$ can be controlled via the parameter λ, called the intensity of $N(t)$.

It is assumed that each vehicle has a characteristic, Z, that can be measured by the sensor to determine whether the vehicle is of type **g** or **b**. In addition, we assume: (1) the attribute of vehicle k is Z_k, where the Z_k, $k = 1, 2, \ldots, N(t)$, are independent copies of random variable Z, and (2) Z can be modeled as a mixture of two distributions, each having the form of a beta random variable. The PDF of Z is

$$f_Z(z; p, \mu) = \begin{cases} pf(z; 0, 1) + (1-p)f(z; \mu + 1, \mu + 3), & 0 \leq z \leq \mu + 3, \\ 0, & \text{otherwise}, \end{cases} \tag{6.5}$$

where $f(\,\cdot\,; \alpha, \beta)$ denotes the PDF of a beta random variable, assumed symmetric on interval $[\alpha, \beta]$, *i.e.*,

$$f(z; \alpha, \beta) = \begin{cases} \dfrac{30}{(\beta - \alpha)^5}(z - \alpha)^2(z - \beta)^2, & \alpha \leq z \leq \beta, \\ 0, & \text{otherwise}, \end{cases} \tag{6.6}$$

$p \in [0, 1]$ denotes the percentage of type **g** vehicles that pass the sensor in $[0, t]$, and $\mu \geq -1$ is a parameter that quantifies the separation between the attributes of type **g** and type **b** vehicles. By Eq. (6.5), the percentage of type **b** vehicles that pass the sensor in $[0, t]$ is $(1 - p)$, the attribute of type **g** vehicles satisfies $0 \leq Z \leq 1$ almost surely (a.s.), and the attribute of type **b** vehicles satisfies $\mu + 1 \leq Z \leq \mu + 3$ a.s. Further, with $\mu < 0$, the attributes of type **g** and type **b** vehicles overlap; for $\mu \geq 0$ they do not. The PDF of Z is plotted in Fig. 6.2 for the case of $p = 3/4$ and $\mu = -1/2$. The generalization of f_Z to the case of three or more vehicle types is straightforward using, for example, higher-order mixtures of beta random variables.

6.1.1.2 Sensor model

Measurement error and/or background noise is present in typical sensor applications [12, 45, 79]. In this section, we construct a mathematical model for the sensor that includes measurement error; the effects of background noise will be discussed in Section 6.1.5.

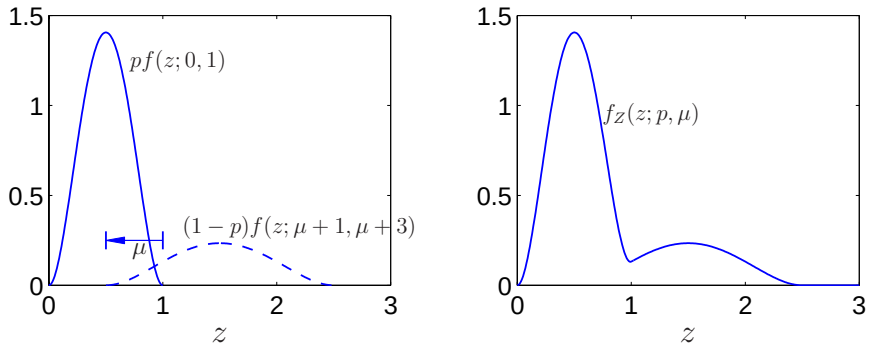

Figure 6.2: PDF of vehicle attribute model, Z.

It is assumed that the output from the sensor can be described via

$$Y(t) = \sum_{k=1}^{N(t)} (Z_k + E_k)\, h(t, T_k; r), \;\; t \geq 0, \tag{6.7}$$

where T_k are the jump times of $N(t)$ in Eq. (6.3), Z_k are the attributes of the vehicles, $\{E_k\}$ is a collection of zero-mean iid Gaussian random variables with variance σ^2 modeling measurement errors, $r \geq 0$ is a deterministic parameter, and

$$h(t, T_k, r) = \begin{cases} 1(t = T_k) & r = 0, \\ \dfrac{1}{r^4}(t + r - T_k)^2 (t - r - T_k)^2 \; 1(t \in [T_k - r, T_k + r]) & r > 0. \end{cases} \tag{6.8}$$

Further, we assume the value for σ^2 can be specified, *i.e.*, used as a design variable; for the special case when $\sigma^2 = 0$, the measurement error is zero. By Eq. (6.8), $h(t)$ is a symmetric pulse, centered at arrival time T_k, with unit height and width $2r$. We use parameter r to model the sensor range. For example, consider a sample of Y with one jump in $[0, t]$. If $r > 0$ and the sensor activates at time T_1, then $Y(t) > 0$, $t \in [T_1 - r, T_1 + r]$; if $r = 0$, the sensor has zero range, meaning that $Y(t) > 0$ for $t = T_1$ only. We consider the special case where $r = 0$ in the analyses that follow; $r > 0$ will be considered in Section 6.1.5.

The sensor classifies vehicles using the following rule:

$$\text{if } \begin{cases} Y(t) \geq \delta & \Rightarrow \text{ a type } \mathbf{b} \text{ vehicle passed at time } t, \\ Y(t) < \delta & \Rightarrow \text{ no type } \mathbf{b} \text{ vehicle passed at time } t, \end{cases} \tag{6.9}$$

where $\delta \geq 0$ is a deterministic parameter that defines the sensitivity of the sensor. With $r = 0$, it is sufficient to evaluate Eq. (6.9) at $Y(T_k)$, $k = 1, 2, \ldots, N(t)$, since $Y(t) = 0$ for $t \notin \{T_1, T_2, \ldots, T_{N(t)}\}$. Note that a misclassification may result if type $\mathbf{b}$ vehicles mask ($\mu < 0$), and/or if measurement error exists ($\sigma^2 > 0$).

One sample of $Y(t)$ from Eq. (6.7) is shown in Fig. 6.3 for the case of $t = 20$, $\mu = -3/4$, $\delta = 3/4$, $\sigma^2 = 1/100$, and $r = 0$; the underlying sample of $N(t)$ exhibits

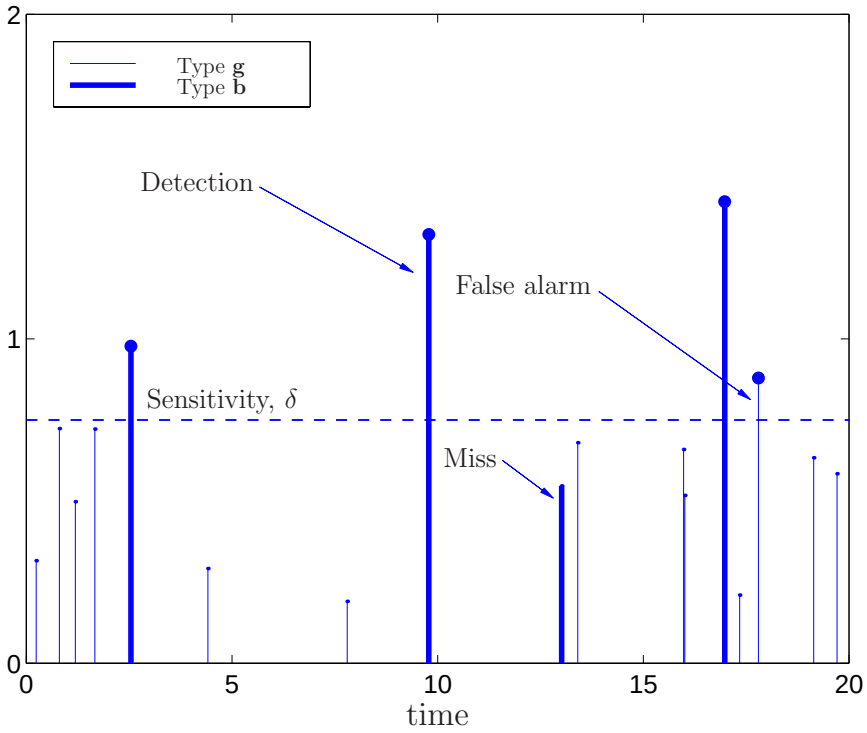

Figure 6.3: One sample of sensor measurement, Y.

17 jumps. Type **g** vehicles are denoted with thin lines and type **b** vehicles with thick lines. The sensitivity, δ, is also shown using a dashed line. By Eq. (6.9), 4 vehicles activate the sensor; one of them is a type **g**. Likewise, 13 vehicles do not activate the sensor; one of them is a type **b**. Hence, for this sample, there are 3 detections and 2 failures: one false alarm and one miss, as defined in Eq. (6.2).

6.1.2 Sensor performance

Metrics are defined to assess the performance of the sensor, and then used as the basis for the design problem. Let $\{W_k\}$, $k = 1, 2, \ldots, N(t)$, be a sequence of random variables such that

$$W_k = \begin{cases} 1 & \text{if } k^{\text{th}} \text{ vehicle is type } \mathbf{b}, \\ 0 & \text{else}, \end{cases} \tag{6.10}$$

and define

$$X(t) = \sum_{k=1}^{N(t)} W_k \; 1(t = T_k). \tag{6.11}$$

The possible outcomes from Table 6.1 can be reformulated in terms of $Y(t)$ in Eq. (6.7) and $X(t)$ in Eq. (6.11) for fixed time, t, as shown in Table 6.2. It follows

Table 6.2. Sensor performance at time, t.

Sensor is	**Type b vehicle**	
	does pass	does not pass
activated	$X(t) = 1$ and $Y(t) \geq \delta$	$X(t) = 0$ and $Y(t) \geq \delta$
not activated	$X(t) = 1$ and $Y(t) < \delta$	$X(t) = 0$ and $Y(t) < \delta$

that the number of detections occurring in $[0, t]$, as defined by Eq. (6.1), can be written as

$$\begin{aligned} D(t) &= \int_0^t 1\left[(X(\tau) = 1) \cap (Y(\tau) \geq \delta)\right] \mathrm{d}\tau \\ &= \begin{cases} 0 & \text{if } N(t) = 0, \\ \sum\limits_{k=1}^{N(t)} 1[(W_k = 1) \cap (Y(T_k) \geq \delta)] & \text{if } N(t) > 0, \end{cases} \end{aligned} \tag{6.12}$$

where the integral reduces to a discrete summation since $r = 0$ in Eq. (6.8). The two types of failures defined in Eq. (6.2), *i.e.*, the off-diagonal terms from Table 6.2, can be written as

$$\begin{aligned} A(t) &= \int_0^t 1\left[(X(\tau) = 0) \cap (Y(\tau) \geq \delta)\right] \mathrm{d}\tau \\ &= \begin{cases} 0 & \text{if } N(t) = 0, \\ \sum\limits_{k=1}^{N(t)} 1[(W_k = 0) \cap (Y(T_k) \geq \delta)] & \text{if } N(t) > 0, \end{cases} \end{aligned} \tag{6.13}$$

and

$$\begin{aligned} M(t) &= \int_0^t 1\left[(X(\tau) = 1) \cap (Y(\tau) < \delta)\right] \mathrm{d}\tau \\ &= \begin{cases} 0 & \text{if } N(t) = 0, \\ \sum\limits_{k=1}^{N(t)} 1[(W_k = 1) \cap (Y(T_k) < \delta)] & \text{if } N(t) > 0, \end{cases} \end{aligned} \tag{6.14}$$

where $A(t)$ and $M(t)$ are the number of false alarms and misses, respectively, occurring in $[0, t]$.

Alternative definitions for Eqs. (6.12) - (6.14) can be used. Define the number of type **b** and type **g** vehicles in $[0, t]$ to be

$$\begin{aligned} N_b(t) &= \sum_{k=1}^{N(t)} W_k, \text{ and} \\ N_g(t) &= N(t) - N_b(t), \end{aligned} \tag{6.15}$$

respectively. Then

$$\bar{D}(t) = \begin{cases} 1 & \text{if } N(t) = 0, \\ 1 & \text{if } N(t) > 0 \text{ and } N_b(t) = 0, \\ \dfrac{D(t)}{N_b(t)} & \text{if } N(t) > 0 \text{ and } N_b(t) > 0, \end{cases}$$

$$\bar{A}(t) = \begin{cases} 0 & \text{if } N(t) = 0, \\ 0 & \text{if } N(t) > 0 \text{ and } N_g(t) = 0, \\ \dfrac{A(t)}{N_g(t)} & \text{if } N(t) > 0 \text{ and } N_g(t) > 0, \text{ and} \end{cases}$$

$$\bar{M}(t) = \begin{cases} 0 & \text{if } N(t) = 0, \\ 0 & \text{if } N(t) > 0 \text{ and } N_b(t) = 0, \\ \dfrac{M(t)}{N_b(t)} & \text{if } N(t) > 0 \text{ and } N_b(t) > 0, \end{cases} \tag{6.16}$$

are the corresponding number of detections, false alarms, and misses occurring in $[0, t]$, normalized by $N_g(t)$ or $N_b(t)$. Note that $\bar{D}(t)$, $\bar{A}(t)$, and $\bar{M}(t)$ take values in $[0, 1]$; we refer to $\bar{D}(t)$, $\bar{A}(t)$, and $\bar{M}(t)$ as the **detection rate**, **false alarm rate**, and **miss rate**, respectively.

Consider a fixed time, t, and let $\bar{A}$ and $\bar{M}$ denote $\bar{A}(t)$ and $\bar{M}(t)$. Let d_1 and d_2 be the maximum allowable false alarm and miss rates, respectively, in $[0, t]$. The design objective for the sensor is given by

$$P(\bar{A} \leq d_1) = \bar{q}_1, \text{ and} \tag{6.17a}$$

$$P(\bar{M} \leq d_2) = \bar{q}_2, \tag{6.17b}$$

where $\bar{q}_1$ and $\bar{q}_2$ are required reliabilities, and $0 \leq d_1, d_2, \bar{q}_1, \bar{q}_2 \leq 1$. If both requirements from Eq. (6.17) cannot be met, we assume the requirement on $\bar{M}$ is more important that the requirement on $\bar{A}$. The design of the sensor does not depend on the detection rate, $\bar{D}(t)$; a third requirement can be added to Eq. (6.17) if necessary.

Table 6.3 lists the parameters used in the analyses that follow. Two cases will be considered and discussed in Sections 6.1.3 and 6.1.4, respectively: (1) $\sigma^2 = 0$, and (2) $\sigma^2 > 0$. For Case #1, measurement error is zero and the sensitivity of the sensor, δ, is the design variable; we select its value so as to satisfy Eq. (6.17). The value for μ is unknown; we cannot know how type **b** vehicles will mask. The design of the sensor under nonzero measurement error is considered in Case #2, where both δ and the variance of the measurement error, σ^2, are design variables, while μ is the unknown. These cases are summarized in Table 6.4.

Table 6.3. Parameters used in detection and classification problem.

Description	Parameter	Value
Traffic intensity	λ	1
Pulse width	r	0
Maximum allowable false alarm rate	d_1	$p/2$
Maximum allowable miss rate	d_2	$(1-p)/2$
Proportion of type **g** vehicles	p	0.9
Required reliability for $\bar{A}$	$\bar{q}_1$	0.9
Required reliability for $\bar{M}$	$\bar{q}_2$	0.9
Length of time interval	t	20

Table 6.4. Design cases for vehicle detection and classification.

	Section	Design variables	Unknown
Case #1	6.1.3	δ	μ
Case #2	6.1.4	δ, σ^2	μ

6.1.3 Case #1: model selection for optimal sensitivity

In this section, we employ the decision-theoretic method for model selection developed in Chapter 2 to choose the optimal sensitivity level of the sensor, δ, for the case of zero measurement error, $\sigma^2 = 0$. The model use is design, so the utility formulation of Section 2.3.2.2 applies.

6.1.3.1 Sensor design

The top-half of Fig. 6.4 shows estimates of the probability laws for $\bar{A}$ (left) and $\bar{M}$ (right), for different values of both μ and δ, using 1,000 samples. We assume μ takes values in $[-1, 0)$. This is sufficient because for $\mu \geq 0$, the attributes of type **b** and type **g** vehicles do not overlap (see Fig. 6.2) and, with $\delta = 1$, $P(\bar{A} \leq d_1) = P(\bar{M} \leq d_2) = 0$ for any $d_1, d_2 \in [0, 1]$, *i.e.*, there is zero probability of a false alarm or miss in $[0, t]$. Note that the false alarm rate, $\bar{A}$, is independent of the masking parameter, μ, since changing μ only effects the attribute of type **b** vehicles.

The top-half of Fig. 6.4 is a surface plot of $P(\bar{A} \leq d_1)$ and $P(\bar{M} \leq d_2)$ as a function of μ and δ; the bottom-half of Fig. 6.4 shows the curves defined by the intersection of

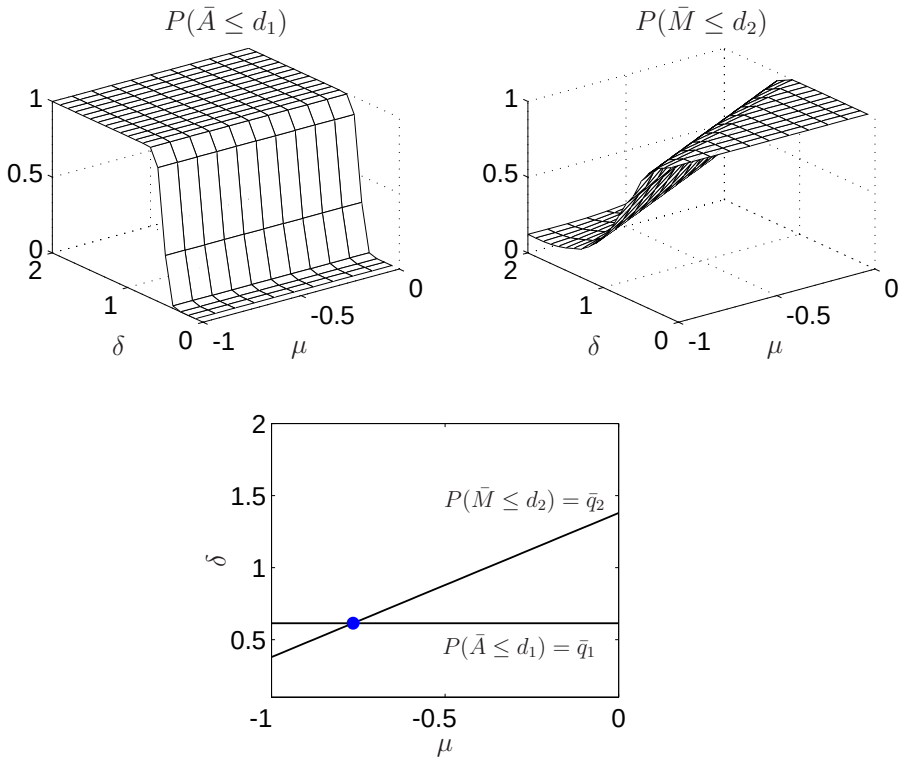

Figure 6.4: Sensor performance for different values of μ and δ.

these two surfaces with the planes $P(\bar{A} \leq d_1) = \bar{q}_1$ and $P(\bar{M} \leq d_2) = \bar{q}_2$, respectively, where it is evident that both requirements of Eq. (6.17) can be satisfied only at a single pair of (μ, δ), denoted by the solid circle. For arbitrary μ, both design requirements cannot be met. This makes intuitive sense because we have two design constraints, but one design parameter, δ.

Assuming a false alarm is preferable to a miss, we design the sensor to satisfy the requirement on $\bar{M}$ alone, Eq. (6.17b). The design equation is then

$$\delta = \mu + 1.379, \tag{6.18}$$

which is the equation to the line $P(\bar{M} \leq d_2) = \bar{q}_2$ shown at the bottom of Fig. 6.4, for the values of d_2 and $\bar{q}_2$ listed in Table 6.3. Equation (6.18) gives the value for the design variable, δ, assuming we know μ. However, we do not know μ perfectly; methods of model selection can be used to select the optimal value for δ.

6.1.3.2 Candidate designs

The possible states of nature are $\mathcal{M} = \{\mu : \mu \in [-1, 0)\}$, the values of the unknown masking parameter. This collection is uncountable; we instead consider a finite sub-collection, $\mathcal{M}' \subset \mathcal{M}$, where

$$\mathcal{M}' = \{m_i\} = \left\{ \mu_i \colon \mu_i = \frac{i-1}{q} - 1 \right\}, \; i = 1, 2, \ldots, q. \tag{6.19}$$

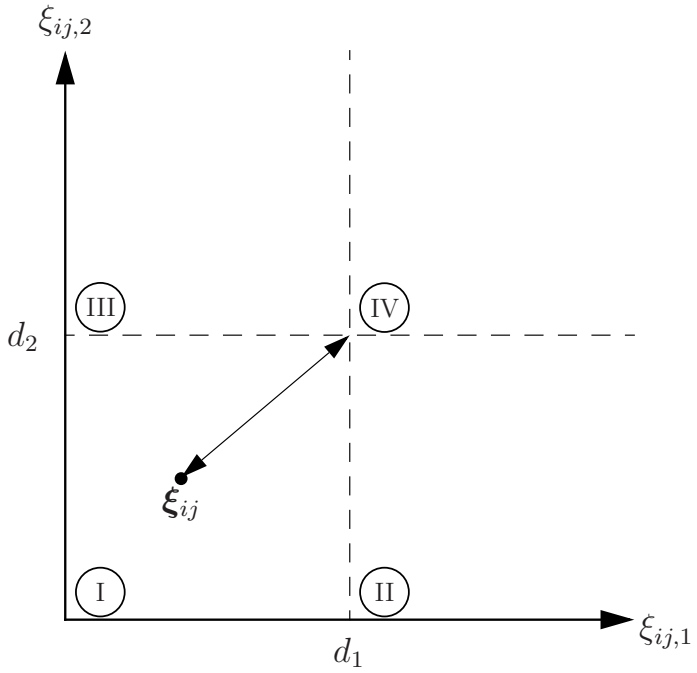

Figure 6.5: The distance $\|\boldsymbol{\xi}_{ij} - \mathbf{d}\|$ used for sensor design.

We use $q = 10$ for calculations, so $\mathcal{M}' = \{-1, -0.9, \ldots, -0.1\}$; as q increases, $\mathcal{M}'$ is an improved approximation for $\mathcal{M}$. The space of candidate designs is

$$\mathcal{A} = \{a_i\} = \{\delta_i \colon \delta_i = \mu_i + 1.379\},\ i = 1, 2, \ldots, q, \tag{6.20}$$

which are the values for sensor sensitivity, δ, required to achieve Eq. (6.17b), assuming $\mu = \mu_i \in \mathcal{M}'$.

Consider the case of design a_i under state of nature m_j. Because the state of nature is unknown, in general the action taken will be incorrect, *i.e.*, for $i \neq j$

$$\begin{aligned} P(\bar{A} \leq d_1 \mid a_i, m_j) &\neq \bar{q}_1, \text{ and} \\ P(\bar{M} \leq d_2 \mid a_i, m_j) &\neq \bar{q}_2. \end{aligned} \tag{6.21}$$

This is further illustrated by Fig. 6.5, which shows the two-dimensional space, $\boldsymbol{\xi}_{ij} = (\xi_{ij,1}, \xi_{ij,2})^T$, where the coordinates of $\boldsymbol{\xi}_{ij}$ are defined via

$$\begin{aligned} P(\bar{A} \leq \xi_{ij,1} \mid a_i, m_j) &= \bar{q}_1, \text{ and} \\ P(\bar{M} \leq \xi_{ij,2} \mid a_i, m_j) &= \bar{q}_2. \end{aligned} \tag{6.22}$$

Also plotted are the lines $\xi_{ij,1} = d_1$ and $\xi_{ij,2} = d_2$; their intersection is the point that achieves both design objectives, $\boldsymbol{\xi}_{ij} = \mathbf{d} = (d_1, d_2)^T$.

6.1.3.3 Optimal design

The design objectives are not satisfied for any arbitrary point, $\boldsymbol{\xi}_{ij}$, that is not coincident with $\mathbf{d}$. Four scenarios are possible, denoted by the four quadrants in Fig. 6.5: (I) the design is conservative with respect to both Eqs. (6.17a) and (6.17b), (II) the design is non-conservative with respect to Eq. (6.17a), but conservative with respect to Eq. (6.17b), (III) the design is conservative with respect to Eq. (6.17a), but non-conservative with respect to Eq. (6.17b), and (IV) the design is non-conservative with respect to both Eqs. (6.17a) and (6.17b). The degree to which the design objectives of Eq. (6.17) are not satisfied can be quantified by

$$\|\boldsymbol{\xi}_{ij} - \mathbf{d}\|^2 = (\xi_{ij,1} - d_1)^2 + (\xi_{ij,2} - d_2)^2, \tag{6.23}$$

the square of the distance from $\boldsymbol{\xi}_{ij}$ to $\mathbf{d}$.

Recall the formulation for the expected utility of design $a_i \in \mathcal{A}$ from Section 2.3.2,

$$u(a_i) = \gamma(a_i) + \sum_{j=1}^{q} \psi(a_i, m_j) p_j, \;\; i = 1, 2, \ldots, q, \tag{6.24}$$

where $\gamma(a_i)$ is the cost of design $a_i \in \mathcal{A}$, $\psi(a_i, m_j)$ is the penalty associated with design a_i, if $m_j \in \mathcal{M}'$ is true, and p_j denotes the probability that m_j is true. We define zero cost for each design and assume each state of nature is equally likely, *i.e.*, $\gamma(a_i) = 0$, $i = 1, 2, \ldots, q$, and $p_j = 1/q$, $j = 1, 2, \ldots, q$. The penalty function, consistent with Section 2.3.2.2, is given by

$$\psi(a_i, m_j) = \tilde{\psi}(\boldsymbol{\xi}_{ij}, \mathbf{d}) = s(\boldsymbol{\xi}_{ij}, \mathbf{d}) \; \|\boldsymbol{\xi}_{ij} - \mathbf{d}\|^2, \tag{6.25}$$

where

$$s(\boldsymbol{\xi}_{ij}, \mathbf{d}) = \begin{cases} 1 & \text{if } \xi_{ij,1} \leq d_1 \text{ and } \xi_{ij,2} \leq d_2, \\ 10 + 90\beta & \text{if } \xi_{ij,1} > d_1 \text{ and } \xi_{ij,2} \leq d_2, \\ 100 - 90\beta & \text{if } \xi_{ij,1} \leq d_1 \text{ and } \xi_{ij,2} > d_2, \\ 110 & \text{if } \xi_{ij,1} > d_1 \text{ and } \xi_{ij,2} > d_2, \end{cases} \tag{6.26}$$

and $\beta \in [0, 1]$ is a deterministic parameter. Note that $s = 1$ and $s = 110$ correspond to the cases where the design is conservative and non-conservative with respect to both design objectives, *i.e.*, quadrants I and IV in Fig. 6.5, respectively. The remaining two cases are a function of β, which can be used to adjust the relative penalty for failure to satisfy the requirements on $\bar{A}$ or $\bar{M}$. For example, with $\beta = 0$ failure to meet the requirements on $\bar{M}$ is penalized ten times greater than failure to meet the requirements on $\bar{A}$; with $\beta = 1$, the reverse is true.

The expected utilities of each design, a_i, $i = 1, 2, \ldots, q$, are computed using Eq. (6.24); the optimal design minimizes Eq, (6.24). Results are shown in Fig. 6.6 as a function of β. The optimal design as a function of β is illustrated with a thick

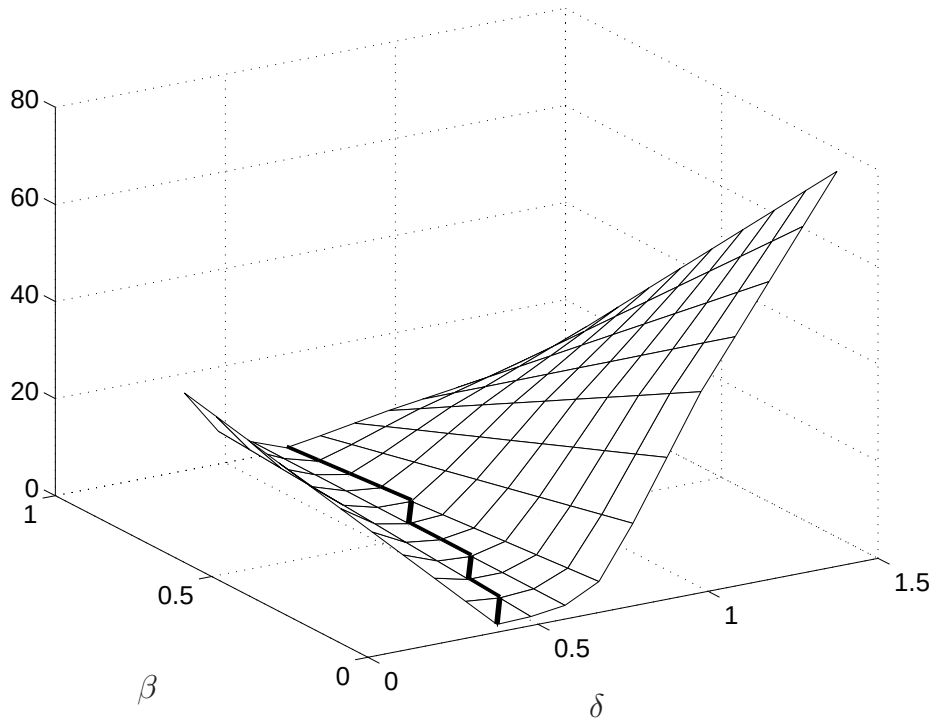

Figure 6.6: Expected utilities, $u(a_i;\beta)$, as a function of β.

line. For β near zero, the formulation heavily penalizes designs that don't satisfy the requirements on the miss rate, $\bar{M}$. Hence it is preferable to have a smaller value of δ to avoid potential misses. As $\beta \to 1$, designs that result in a high false alarm rate are highly penalized; it is therefore preferable to have a larger value of δ to minimize potential false alarms. With β near zero, the optimal sensitivity is $\delta = 0.38$; as β approaches one, the optimal sensitivity increases to $\delta = 0.68$.

6.1.4 Case #2: model selection for optimal sensitivity and accuracy

In this section, we employ model selection techniques to select the optimal design of the sensor in the presence of measurement error. The probability law for the detection rate of the sensor is also calculated.

6.1.4.1 Sensor design

To design the sensor in the presence of measurement error, we must specify values for: (1) the sensitivity, δ, and (2) the variance of the measurement error, σ^2. Recall the model for sensor measurement error, $\{E_k\}$, a sequence of iid random variables, where E_1 is zero-mean, Gaussian with variance σ^2. Consider the case where there

Table 6.5. The accuracy of five sensor types considered for design.

Sensor type	y	σ^2
1	0.05	4.62×10^{-4}
2	0.1	1.85×10^{-3}
3	0.5	4.62×10^{-2}
4	1	0.185
5	1.5	0.416

are five types of sensors available for use, each with a different, but known, level of accuracy; the accuracy of each sensor type is quantified via

$$P(E_1 \leq y) = 0.99. \tag{6.27}$$

Five values for y are listed in Table 6.5, corresponding to the accuracy of the five sensor types. The variance of E_1 for each sensor type, given by

$$\sigma^2 = \left[\frac{y}{\Phi^{-1}(0.99)}\right]^2 \tag{6.28}$$

where Φ denotes the standard Gaussian CDF, is also listed. By Eqs. (6.27) and (6.28), the type #1 sensor is the most accurate since 4.62×10^{-4} is the smallest of the values for σ^2 considered, while the type #5 sensor is the least accurate since 0.416 is the largest. The corresponding PDFs of the five candidate models considered for E_1 are shown in Fig. 6.7.

Figure 6.8 shows values for μ and δ that satisfy Eq. (6.17) for fixed σ^2. We assume μ takes values in $[-1, 0.2]$; unlike Case #1, there is a nonzero probability of a false alarm or a miss in $[0, t]$ for $\mu \geq 0$ due to the presence of measurement error. The horizontal lines in Fig. 6.8 define the μ and δ that satisfy Eq. (6.17a) for fixed σ^2; the lines with nonzero slope define the μ and δ that satisfy Eq. (6.17b) for fixed σ^2. Five different line types are plotted, which correspond to the five values for σ^2. A solid circle is used to indicate the triple (μ, δ, σ^2) where both design objectives can be satisfied simultaneously; there are five such triples. Hence, unlike the design in Section 6.1.3.1, there are multiple points in the design space where Eq. (6.17) is satisfied; the values for δ and σ^2 at these points define the candidate sensor designs.

6.1.4.2 Candidate designs

The possible states of nature are $\mathcal{M} = \{\mu : \mu \in [-1, 0.2]\}$, the values of the unknown masking parameter. This collection is uncountable; we instead consider a

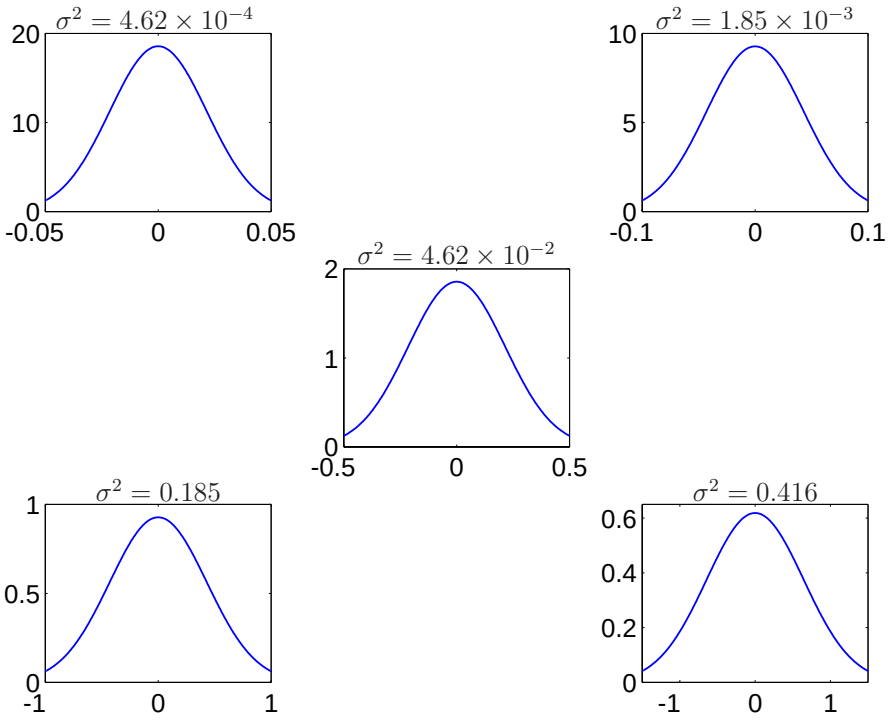

Figure 6.7: Candidate models for E_1.

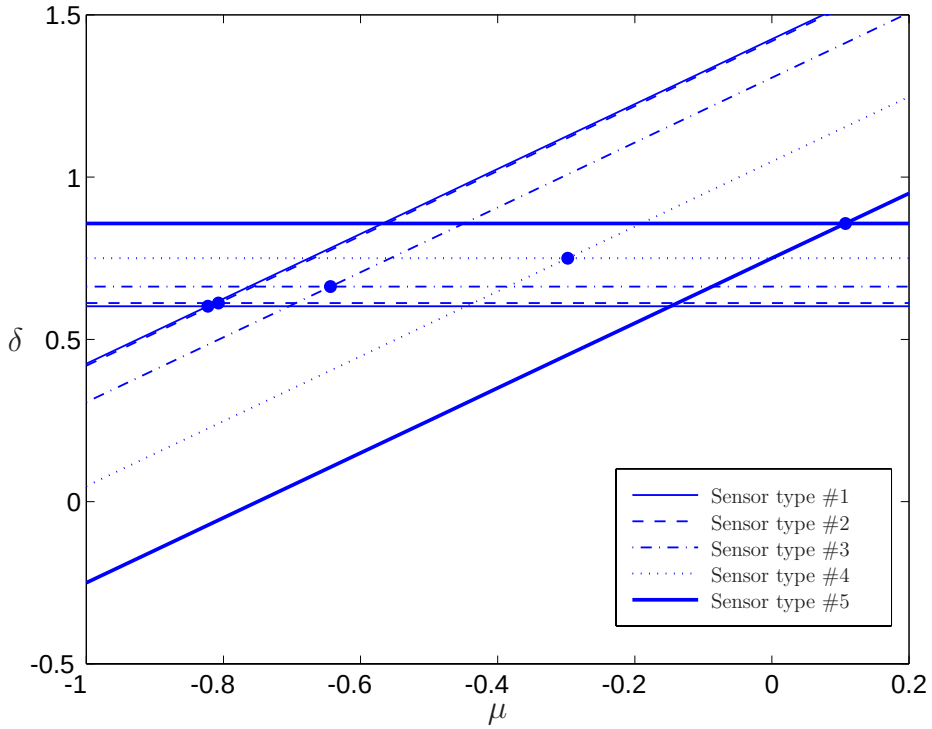

Figure 6.8: Design space for case #2.

Table 6.6. Candidate designs for case #2.

Design	δ	σ^2
a_1	0.602	4.62×10^{-4}
a_2	0.612	1.85×10^{-3}
a_3	0.662	4.62×10^{-2}
a_4	0.750	0.185
a_5	0.857	0.416

finite subcollection, $\mathcal{M}' \subset \mathcal{M}$, where

$$\mathcal{M}' = \{m_i\} = \left\{ \mu_i \colon \mu_i = \frac{6(i-1)}{5(q-1)} - 1 \right\}, \ i = 1, 2, \ldots, q. \tag{6.29}$$

We use $q = 13$ for calculations, so $\mathcal{M}' = \{-1, -0.9, \ldots, 0.2\}$. The space of candidate designs is

$$\mathcal{A} = \{a_i\} = \{(\delta_i, \sigma_i^2)\}, \ i = 1, 2, \ldots, 5, \tag{6.30}$$

which are the values for δ and σ^2 from the five triples shown in Fig. 6.8; the elements of $\mathcal{A}$ are listed in Table 6.6.

6.1.4.3 Optimal design

The procedure to select the optimal $a_i \in \mathcal{A}$ is identical to the case of zero measurement error, *i.e.*, Section 6.1.3.3, with one exception: the cost for each candidate design is no longer zero. Instead, define

$$\gamma(a_i) = \tilde{\gamma}(\sigma_i) = \frac{\gamma_0}{\sigma_i^2}, \ i = 1, 2, 3, 4, 5, \tag{6.31}$$

where $\gamma_0 > 0$ is a deterministic parameter, $\sigma_i^2 > 0$ denotes the variance of the measurement error under design a_i, *i.e.*, row i from Table 6.6, and we replace γ with $\tilde{\gamma}$ to denote that the cost of design a_i can be expressed as a function of σ_i alone. By Eq. (6.31), designs that require a very accurate sensor (small σ^2) have a high cost.

The expected utility for each $a_i \in \mathcal{A}$ is shown in Fig. 6.9; parameter $\beta = 1$ in Eq. (6.26) was used for calculations. Three lines are plotted, corresponding to three values for γ_0; in each case, the optimal design, denoted by $a^\star$, is indicated with a solid circle. With $\gamma_0 = 1/100$, design $a_3 = (\delta_3, \sigma_3^2) = (0.662, 4.62 \times 10^{-2})$ is optimal. For small γ_0, accurate sensors (small σ^2) have little cost when compared to inaccurate sensors (large σ^2), so the more accurate sensor is optimal. As $\gamma_0 \to 0$, the results from Section 6.1.3.3 are recovered. For $\gamma_0 = 1$, design $a_5 = (\delta_5, \sigma_5^2) = (0.857, 0.416)$ is optimal. In this case, accurate sensors have considerable cost, and a less accurate sensor is optimal; the optimal value for δ is significantly larger than for the case with zero measurement error.

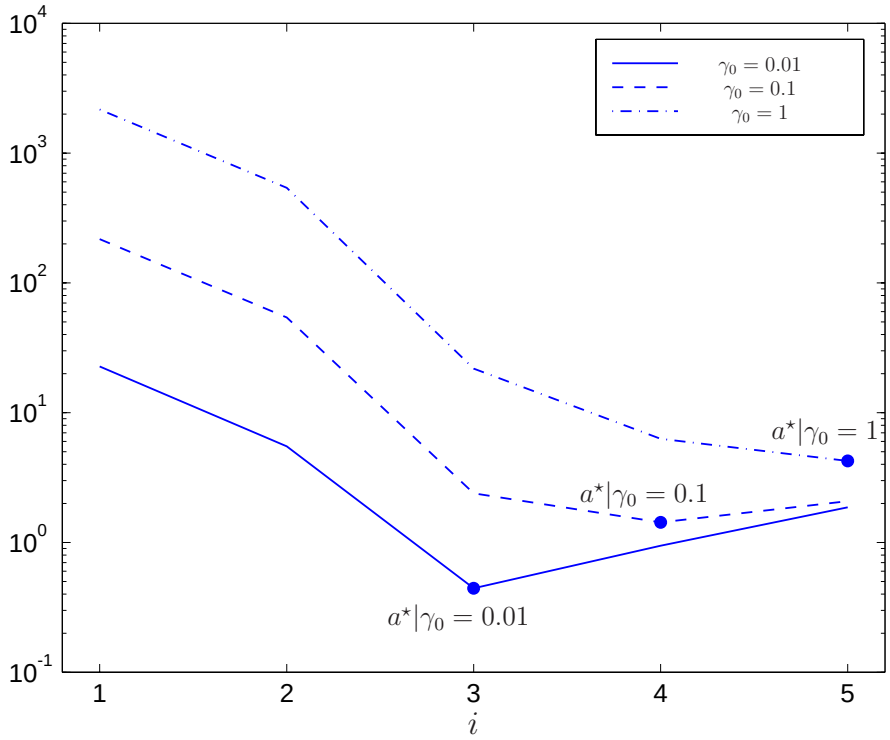

Figure 6.9: Expected utility of each $a_i \in \mathcal{A}$ for case #2.

6.1.4.4 Probability of detection

The probability law for the detection rate of the sensor is a third metric for performance defined in Section 6.1.2; it was not included in the design process, but may be useful to study. Consider a fixed time, t, and let $\bar{D}$ denote $\bar{D}(t)$ from Eq. (6.12). Estimates of $P(\bar{D} > v)$ using 1,000 samples are shown in Fig. 6.10 for the case of $v = 0.9$. The plot shows values for the two design variables, δ and σ^2, such that the probability of detection is greater than 90% in $[0, t]$, for four values of the unknown μ. In all cases, as δ increases, $P(\bar{D} > v) \to 0$, and as σ^2 increases, $P(\bar{D} > v)$ decreases. Further, for fixed μ, the probability of detection is a maximum for $\delta = \sigma^2 = 0$; this maximum is one for $\mu > -1$. The results for zero measurement error are denoted by a dark line; this corresponds to the probability of detection for the sensor design considered in Section 6.1.3.

6.1.5 Sensor design with background noise

In the presence of background noise, we assume: (1) the output of the sensor is given by

$$\tilde{Y}(t) = Y(t) + S(t), \ t \geq 0, \tag{6.32}$$

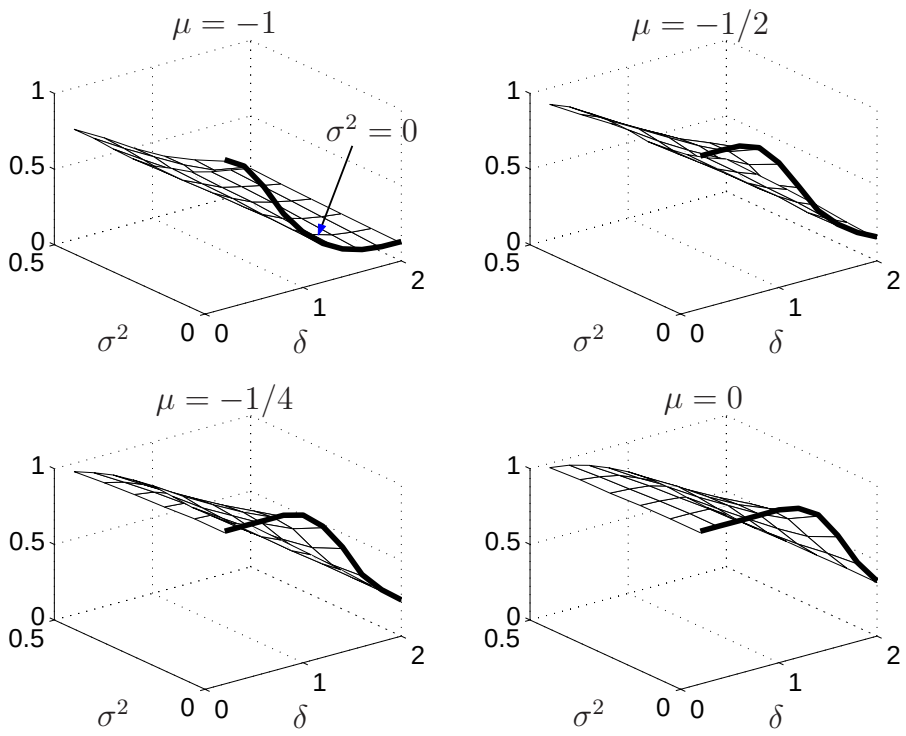

Figure 6.10: Probability of detection estimates, $P(\bar{D} > v)$.

where $Y(t)$ is given by Eq. (6.7), and $S(t)$ is a continuous-time, stationary stochastic process modeling background noise, (2) $S(t) > 0$ a.s., $\forall t \geq 0$, and (3) the magnitude of the noise is small when compared to the magnitude of the signal, *i.e.*, $S(t) \ll Y(t)$ a.s., $\forall t \geq 0$. By (2), we have $\tilde{Y}(t) > Y(t)$ a.s., $\forall t \geq 0$, meaning that when compared to the case with zero background noise there is, for fixed δ and σ^2: (a) no change or an increase in the number of detections, $D(t)$, (b) no change or an increase in the number of false alarms, $A(t)$, and (c) no change or a decrease in the number of misses, $M(t)$. Assumption (3) implies good signal-to-noise ratio; if instead $S \sim Y$, it will be impossible to achieve a satisfactory design of the sensor.

Because S is a continuous-time process, we set $r > 0$ in Eq. (6.8); if $r = 0$, Y takes a finite number of nonzero values in $[0, t]$, and the events defined by Eqs. (6.1) and (6.2) have zero measure. Figure 6.11 shows $Y(t)$, $t \in [0, 3]$, for $r = 0$ (left) and $r = 1/4$ (right). Note that with $r = 1/4$, Y is nonzero everywhere between the second and third peaks; for larger r this effect is amplified, making it difficult or impossible to distinguish between two or more vehicles. Small, nonzero, values for r are therefore optimal.

For calculations, $\tilde{Y}(t)$ is treated as a discrete-time stochastic process. Consider the partition $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = t$ of $[0, t]$; Eq. (6.32) can be written for every point in the partition, *i.e.*,

$$\tilde{Y}(t_j) = Y(t_j) + S(t_j), \ j = 0, 1, \ldots, n. \tag{6.33}$$

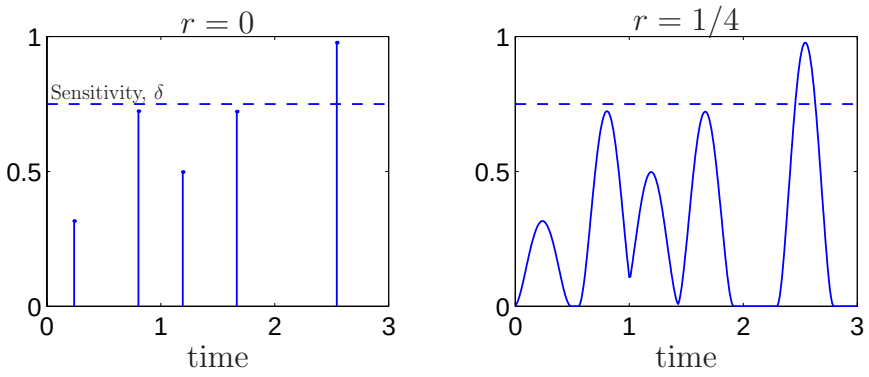

Figure 6.11: Sensor model, Y, with $r = 0$ (left) and $r > 0$ (right).

Define

$$\rho = \frac{1}{\min\limits_{j\in\{1,2,\ldots,n\}} (t_j - t_{j-1})} \tag{6.34}$$

to be the sample rate of the sensor. We assume ρ is a controllable parameter; if ρ is not large enough, aliasing will occur, which may adversely affect the performance of the sensor.

For illustration, consider S to be a translation process, *i.e.*,

$$S(t) = F^{-1} \circ \Phi[G(t)], \ t \geq 0, \tag{6.35}$$

where $G(t)$ is a zero-mean, unit variance, stationary Gaussian white noise process, and F is given by

$$F(x; \alpha, \zeta) = \int_0^x \frac{1}{\sqrt{2\pi}\zeta u} \exp\left[-\frac{1}{2}\left(\frac{\ln u - \alpha}{\zeta}\right)^2\right] \mathrm{d}u, \ x > 0, \tag{6.36}$$

where α and $\zeta > 0$ are deterministic parameters. By Eqs. (6.35) and (6.36), $S(t)$ is a lognormal random variable for any fixed t. One sample of $\tilde{Y}$ is shown in the bottom of Fig. 6.12 for the case of $\alpha = -5$ and $\zeta^2 = 3/2$; the underlying samples of Y and S are shown at the top-left and top-right of Fig. 6.12, respectively. Note that the sample for Y is identical to the one shown in Fig. 6.3, with $r = 1/4$ instead of $r = 0$. By Eq. (6.9), the sensor is activated 6 times: twice due to a type **g** vehicle, three times due to type **b** vehicles, and once more due to background noise. One type **b** vehicle is missed. Hence for this sample, the number of false alarms has increased by two because of the noise; the number of detections and misses remain unchanged.

To design the sensor in the case of nonzero background noise, the formulation of Sections 6.1.3 and 6.1.4 can be used. Parameters r and ρ are design variables and can be included in the space of candidate designs, $\mathcal{A}$. Similarly, the properties of the noise are unknown; the parameters α and ζ, as well as the probability law for S, can be included in $\mathcal{M}$.

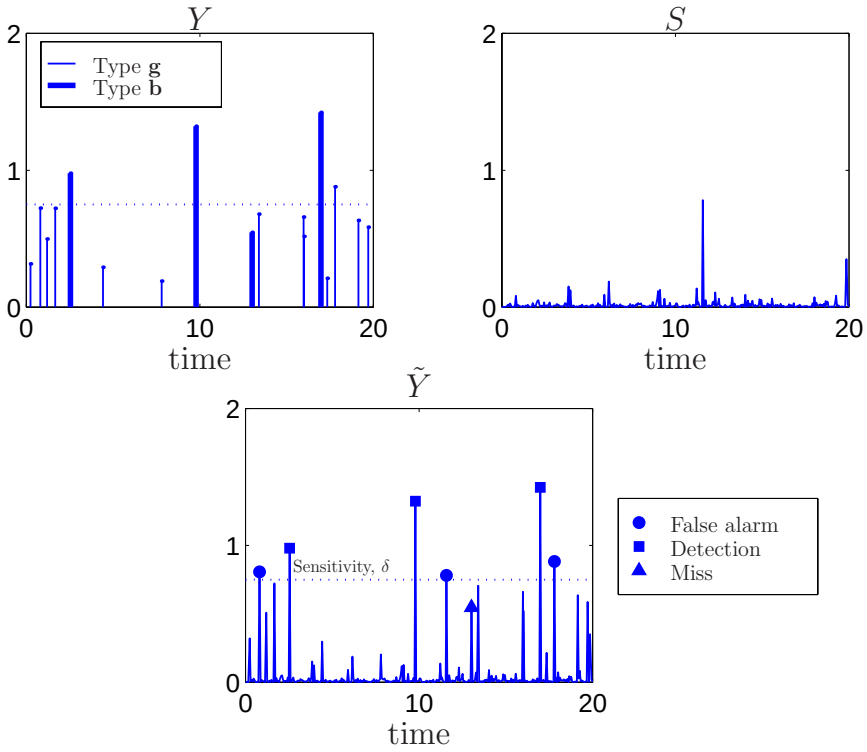

Figure 6.12: One sample of $\tilde{Y}$.

6.2 Vehicle monitoring

Consider the general domain, $\Omega \subseteq \mathbb{R}^2$, with multiple points of attraction, called **treasures**, shown in Fig. 6.13. There are numerous **entry points** into Ω, and multiple **paths** from each entry point to each treasure. We wish to monitor the movement of objects within Ω along all possible paths. Objects may enter Ω at one of the entry points and take one path to a single treasure. We assume a limited number of sensors can be placed at arbitrary locations inside the domain to monitor the objects inside Ω. Each sensor has finite field of view or range, meaning it is possible to monitor only those objects that pass within a disc, centered at the sensor location, with finite radius; the sensor range is denoted by the gray regions in Fig. 6.13. The objective is to design a network of sensors to estimate the path of each object through Ω, under some set of constraints.

One application of this general framework is vehicle surveillance, where the objects of interest are vehicles that are constrained to operate within some network of roads; the objective is to observe the movement of vehicles throughout a given region. The roads form paths from one city to another; the treasures are then a subcollection of these cities. The boundary of the observed region defines Ω for this application, and any road crossing this boundary is an entry point into Ω.

For illustration, we consider Ω to be the road network in New York state, shown in

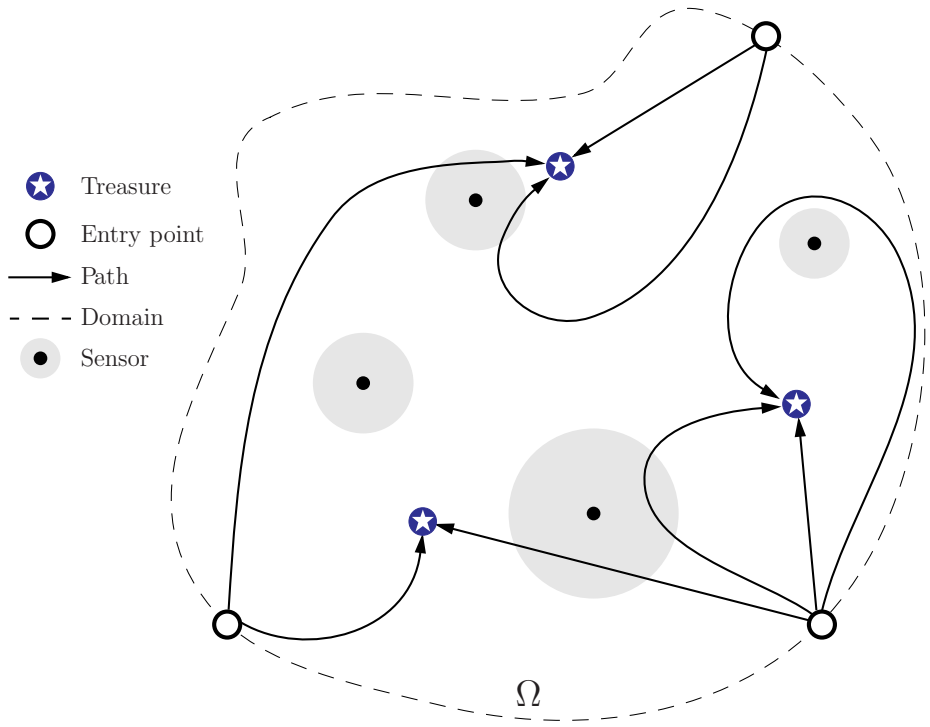

Figure 6.13: General framework for monitoring problem.

Fig. 6.14, where many of the major highways and cities have been included. The level of detail is chosen arbitrarily; additional cities and roads can be added if necessary. We assume a single treasure, the city of Ithaca, and a single type of vehicle, and wish to monitor the movement of all vehicles inside Ω. A limited number of sensors to monitor the passage of a vehicle may be placed directly on a road, or at some location in between roads, where each sensor has a finite range. The objective is to find the optimal placement and range of these sensors, under some set of constraints.

Mathematical models for both the vehicles and sensor network are developed in Section 6.2.1. Metrics that define sensor network performance are developed in Section 6.2.2, and the design of the sensor network is discussed in Section 6.2.3. The sensitivity of the output of interest to certain assumptions on the model of the sensor network is discussed in Section 6.2.4. Methods of model selection, discussed in Section 6.2.5, are then used to compute the optimal design of the sensor network. Possible extensions to this work are presented in Section 6.2.6.

6.2.1 Model development

In this section, we formulate the mathematical models for: (1) the possible paths through the domain, Ω, (2) the vehicle movement on paths within Ω, and (3) the

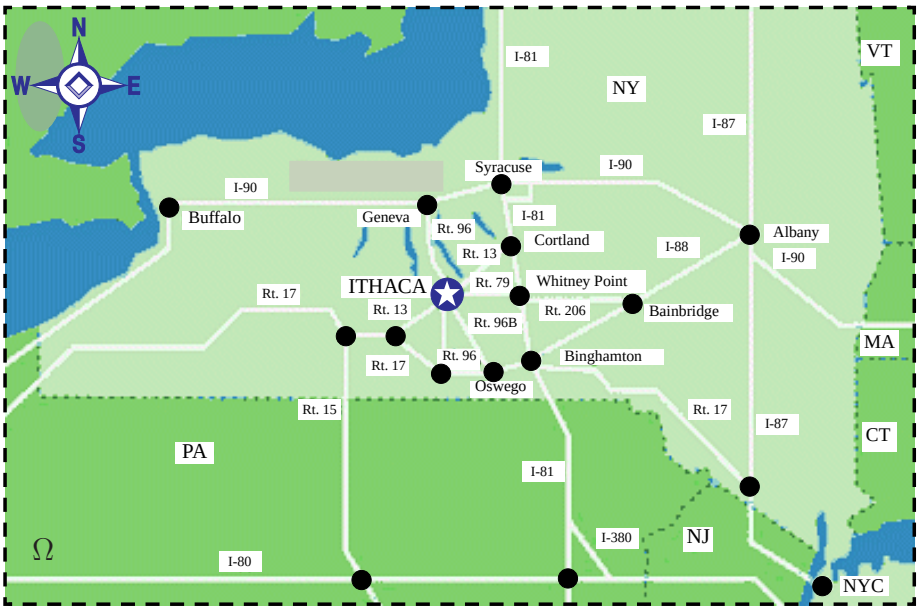

Figure 6.14: Domain of interest: New York state.

sensor network.

6.2.1.1 Model for possible paths

Methods from graph theory [73] are used to formulate a model for paths in Ω. We define the intersection of two or more roads as a **node**, and the line segment connecting any two nodes as an **edge**. A collection of nodes and edges defines a **graph**. The number of edges at node i is called the **degree** of node i. Two nodes connected via an edge are called **adjacent**. The **adjacency matrix**, **c**, is a square matrix used to define the connectivity of the graph; the elements of **c** satisfy

$$c_{ij} = \begin{cases} 1, & \text{if nodes } i \text{ and } j \text{ are adjacent,} \\ 0, & \text{else,} \end{cases} \tag{6.37}$$

where i, j are defined for all integers in $\{1, 2, \ldots,$ number of nodes$\}$. The process of moving between two adjacent nodes is called a **step**. A **path** through Ω is then a sequence of steps; it can be defined via a list of nodes or, equivalently, a list of edges. In the analyses that follow, we will use the node list to define each path.

To illustrate these definitions, consider the graph of Ω shown in Fig. 6.15. The nodes are numbered sequentially from 1 to 21; there are 28 edges in the graph of Ω. The treasure is the city of Ithaca, denoted by the ★ and labeled node #1, while Albany is node #12, Cortland is node #8, Syracuse is node #9, *etc.* There are 8 nodes of degree one in the graph of Ω; these are the entry nodes into the domain. For example, node #18 represents entering Ω from near Buffalo (the west); node #14, from New York City (the south). Further, nodes #11 and #12 are adjacent, while

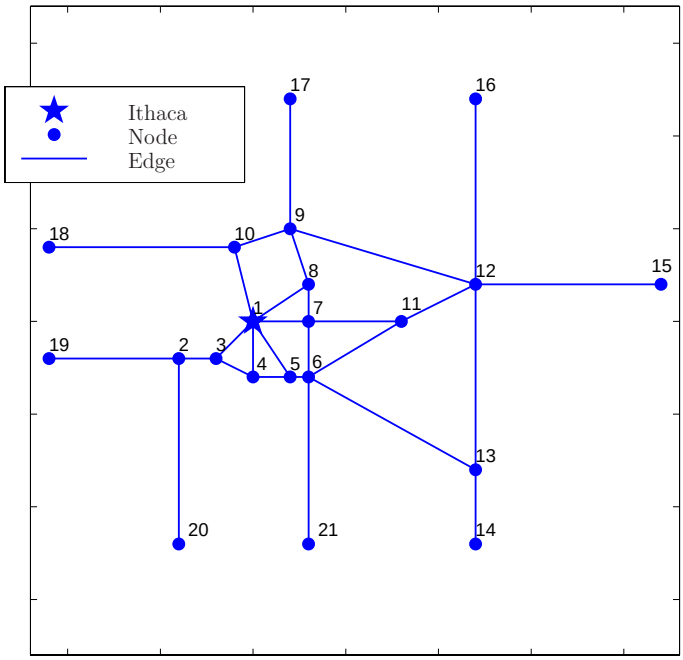

Figure 6.15: Graph of Ω.

nodes #8 and #11 are not. The path from Ithaca to Syracuse by way of Cortland involves 2 steps, and the path is defined as {#1, #8, #9}.

6.2.1.2 Model for vehicle movement

To construct a mathematical model for vehicle movement within Ω, we assume each vehicle travels at constant, known velocity on a path in Ω. In addition, we assume the vehicle can take any path in Ω, provided: (1) the path is nonintersecting, meaning the vehicle can pass over any node only once, (2) the path starts at any of the entry nodes, and ends at the treasure, and (3) the path taken is independent of sensor locations. Under assumptions (1)-(3), the number of paths a vehicle can take through Ω can be calculated; the procedure to do so is discussed in Appendix D. It can be shown that there are 213 paths through Ω. Note that if we relax any or all of assumptions (1)-(3), the number of paths through Ω will change. However, the procedure that follows remains unchanged.

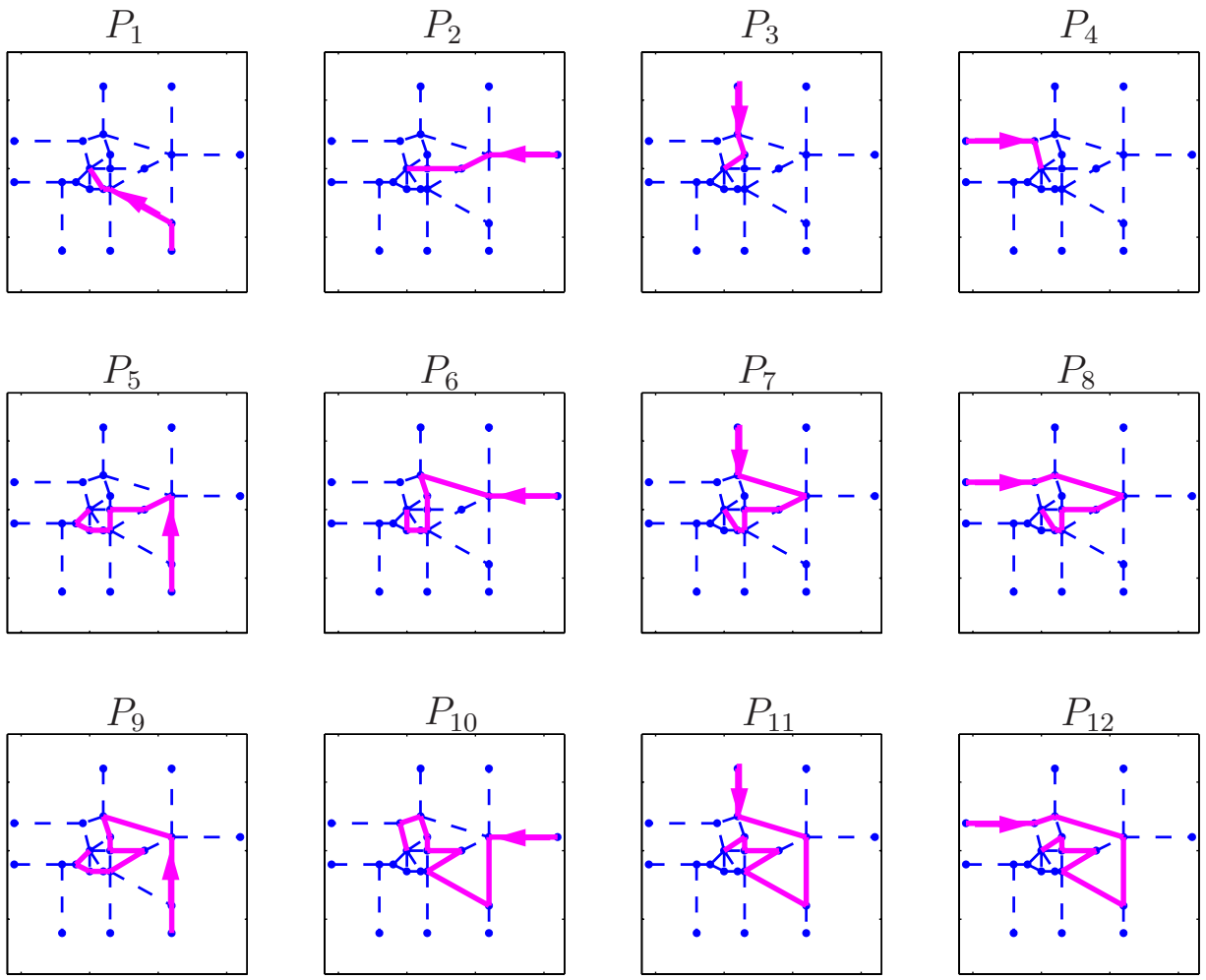

Figure 6.16: The possible paths in Ω, $\{P_j\}$, $j = 1, 2, \ldots, n$.

For calculations, we consider a subset of $n = 12$ of the 213 paths. Only those paths that originate from entry nodes #14, #15, #17, or #18 were retained, which corresponds to paths originating from the south, east, north, or west, respectively. From this collection, those paths with the minimum, maximum, and nearest to average length for each entry node were retained. The subset considered, denoted $\{P_j\}$, $j = 1, 2, \ldots, n$, is called the collection of **prior** or **possible** paths, and is shown in Fig. 6.16. Paths are denoted with solid lines, with arrows to indicate the direction of travel. We denote the actual path of a vehicle by $P^\star$, where $P^\star \in \{P_1, P_2, \ldots, P_n\}$. Note that our choice of $\{P_j\}$ is arbitrary; the subset is used to reduce the amount of calculations required in the analyses that follow.

6.2.1.3 Sensor network model

The mathematical model for the sensor network has three elements: (1) the number of sensors, q, (2) the location of each sensor, and (3) the range of each sensor. We make the following assumptions: (a) q is strictly less than 28, the number of edges in the graph of Ω, (b) the q sensors are arranged in Ω at fixed locations $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_q$, (c) all sensors have the same properties, with adjustable range, $r \in [r_1, r_2]$, where $r_1, r_2 \geq 0$ are deterministic parameters, and (d) if a vehicle passes within range of a sensor, it is detected with probability one. By assumption (a), note that if q is equal to or greater than the number of edges in the graph of Ω, the design of the sensor network is trivial. In this case, we can place one sensor on each edge, and adjust the range small enough so the range of no two sensors overlap; any path $P^\star \in \{P_j\}$ can therefore be identified with probability one. Assumption (a) is not restrictive, since it would be impractical to place a sensor on every road segment in New York state. If we remove assumption (d), the models developed in Section 6.1 for vehicle detection can be included in the analyses.

Four sensor layouts are considered, shown in Figs. 6.17-6.20, respectively. The layouts are defined as follows.

- layout #1: a rectangular grid with $q = 9$ sensors.
- layout #2: $q = 9$ sensors placed on q distinct edges.
- layout #3: a rectangular grid with $q = 12$ sensors.
- layout #4: $q = 12$ sensors placed on q distinct edges.

For layouts #2 and #4, the edges chosen were those most commonly used by the n possible paths.

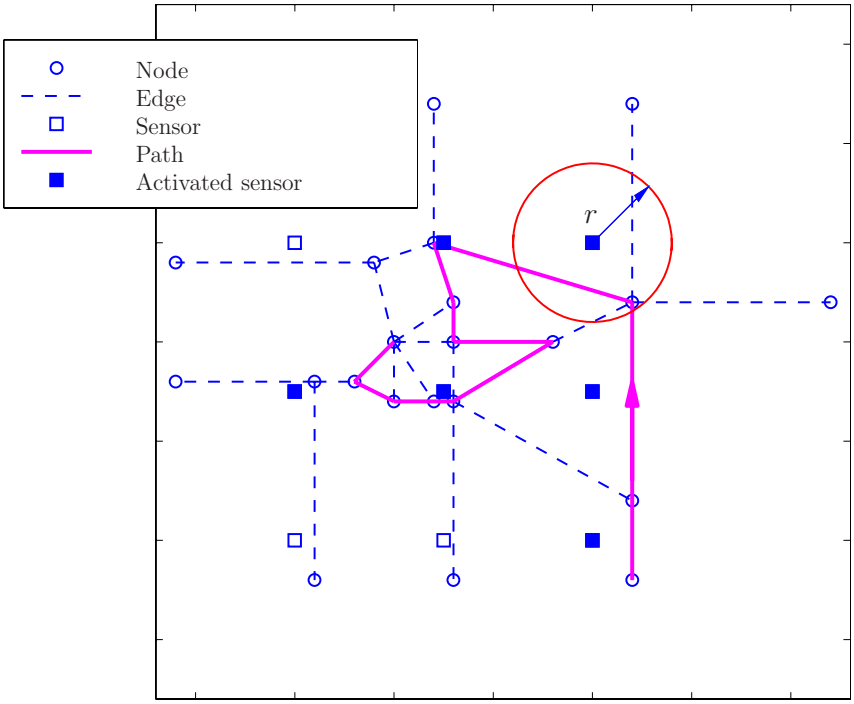

Figure 6.17: Sensor layout #1: $q = 9$; $P^\star = P_9$.

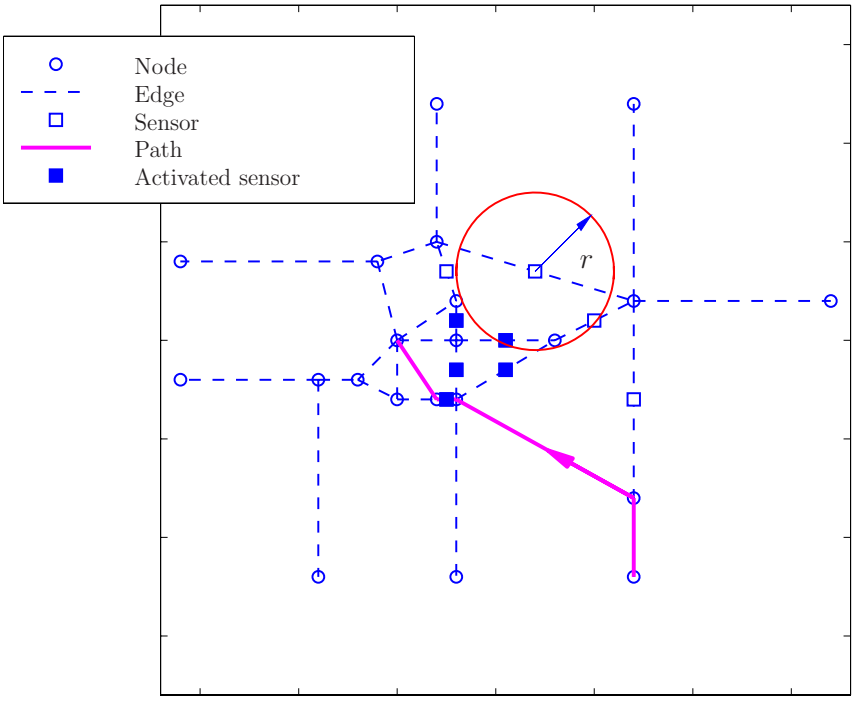

Figure 6.18: Sensor layout #2: $q = 9$; $P^\star = P_1$.

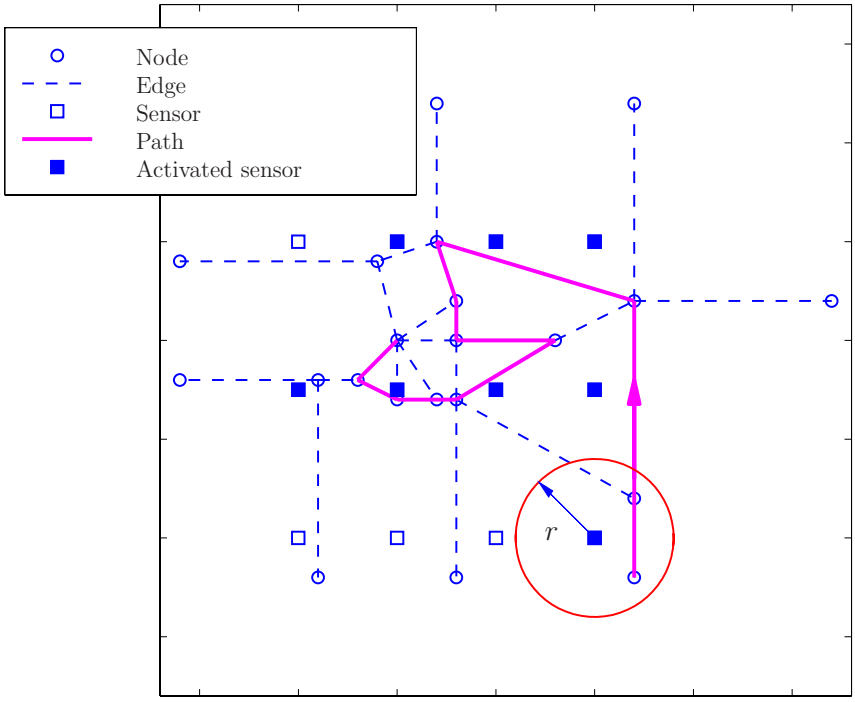

Figure 6.19: Sensor layout #3: $q = 12$; $P^\star = P_9$.

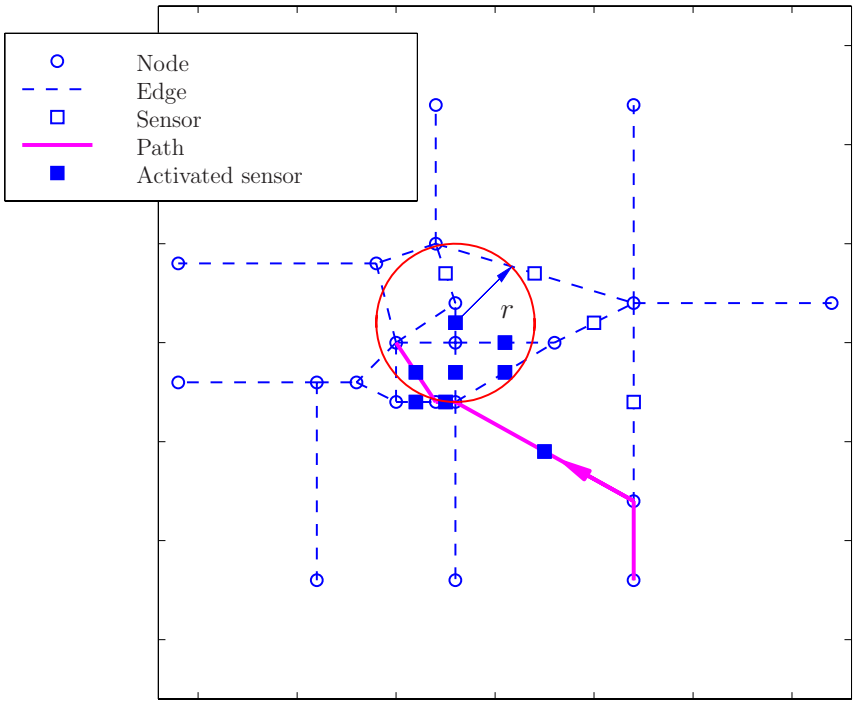

Figure 6.20: Sensor layout #4: $q = 12$; $P^\star = P_1$.

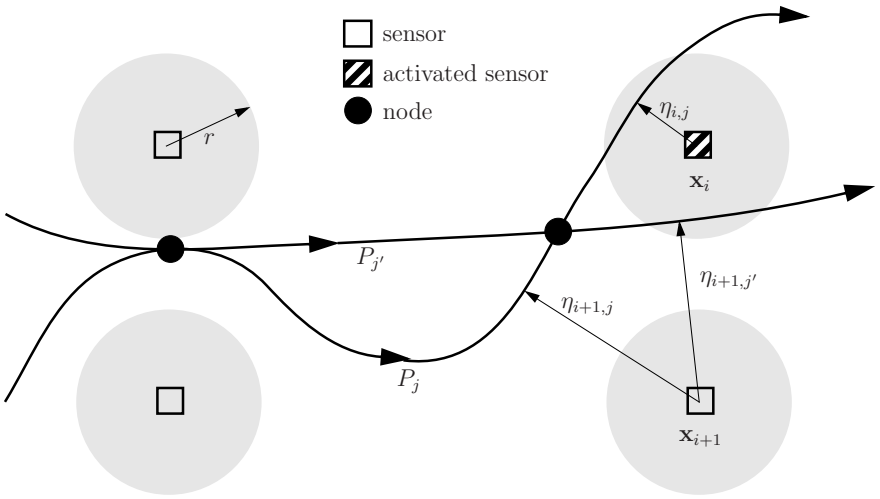

Figure 6.21: The distance $\eta_{i,j}$ defined by Eq. (6.40).

Let $\{w_i\}$, $i = 1, 2, \ldots, q$, be a sequence such that

$$w_i = \begin{cases} 1 & \text{if sensor } i \text{ activates} \\ 0 & \text{else.} \end{cases} \tag{6.38}$$

Note that w_i depends on the range of sensor i, as well as the true path of the vehicle, $P^\star$. Refer again to Figs. 6.17-6.20, where a square denotes the location of a sensor. The symbol ■ denotes activated sensors ($w_i = 1$), while a □ indicates those sensors that were not activated ($w_i = 0$), due to a vehicle taking path $P^\star$. Figures 6.17 and 6.19 are for the case of $P^\star = P_9$, while Figs. 6.18 and 6.20 are for the case $P^\star = P_1$.

The only knowledge of the actual path of the vehicle, $P^\star$, comes from: (1) the sequence $\{w_i\}$, $i = 1, 2, \ldots, q$, and (2) the collection of all possible paths, $\{P_j\}$, $j = 1, 2, \ldots, n$. Accordingly, define the following metric for each path

$$y_j = \frac{1}{q} \sum_{i=1}^{q} 1(\eta_{i,j} < r)\, w_i, \;\; j = 1, 2, \ldots, n, \tag{6.39}$$

where

$$\begin{aligned} \eta_{i,j} = {} & \text{distance between sensor } i \text{ and point of closest approach} \\ & \text{by path } P_j \text{ to sensor } i, \end{aligned} \tag{6.40}$$

is illustrated by Fig. 6.21. For any path P_j, y_j has the following properties: (1) $0 \leq y_j \leq 1$, (2) as $r \to 0$, $y_j \to 0$, and (3) as $r \to \infty$, $y_j \to 1$. In addition, for path $P_j = P^\star$, y_j is a maximum.

Note that Eq. (6.39) implies perfect communication between sensors, *i.e.*, if one sensor detects a vehicle, it is perfectly communicated to all sensors in the network. Models for imperfect communication between sensors is a topic of much research [52]; this issue is not considered here.

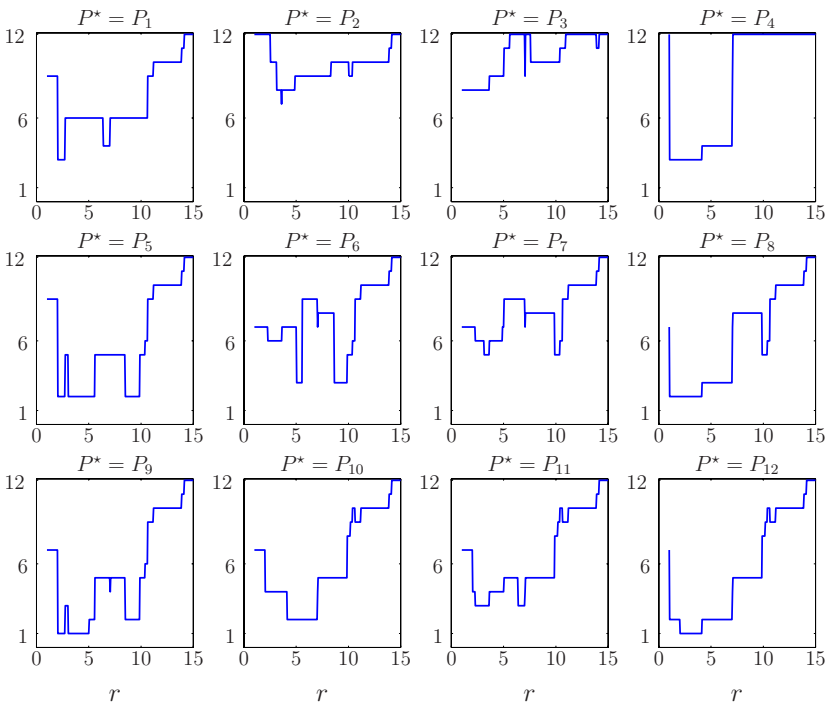

Figure 6.22: The number of posterior paths in Ω, $Q(r)$, for layout #1.

6.2.2 Sensor network performance

Let $\bar{y} = \max_j y_j$, $j = 1, 2, \ldots, n$, and define

$$\tilde{P} = \{P_j \colon y_j = \bar{y}, j = 1, 2, \ldots, n\}. \tag{6.41}$$

The collection $\tilde{P} \subset \{P_1, P_2, \ldots, P_n\}$ defines the set of **posterior paths**, *i.e.*, the collection of paths that cannot be excluded from $\{P_1, P_2, \ldots, P_n\}$ using the sequence $\{w_i\}$, $i = 1, 2, \ldots, q$, the available information from the sensor network. Note that $\tilde{P}$ consists of one or more members of $\{P_1, P_2, \ldots, P_n\}$, in general, and $\tilde{P}$ coincides with $\{P_1, P_2, \ldots, P_n\}$ if $r \to \infty$.

Let

$$Q = \text{the number of members in } \tilde{P}. \tag{6.42}$$

$\tilde{P}$ and Q have the following properties: (1) if $n = 1$, $\tilde{P} = \{P^\star\}$ and $Q = 1$, (2) if $n > 1$, $Q(r) \in \{1, 2, \ldots, n\}$, and (3) for any n, $P^\star \in \tilde{P}$. We see by (2) that the set of posterior paths may not be unique. Q is plotted in Fig. 6.22, as a function of r, for $P^\star = P_j$, $j = 1, 2, \ldots, n$, and sensor layout #1. A similar plot for layout #2 is shown in Fig. 6.23.

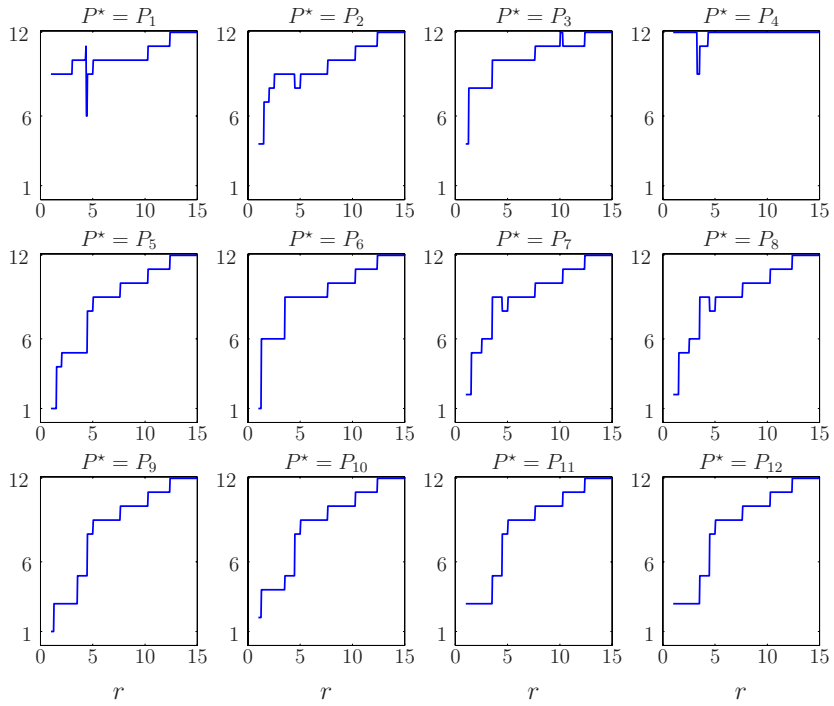

Figure 6.23: The number of posterior paths in Ω, $Q(r)$, for layout #2.

6.2.3 Sensor network design

The design of the sensor network includes: (1) the sensor layout, *i.e.*, the number of sensors and their locations, and (2) the sensor range. It is desirable to design the sensor network so that number of posterior paths in Ω is small, *i.e.*,

$$\text{choose sensor layout and } r \in [r_1, r_2] \text{ such that } Q(r) \leq d, \tag{6.43}$$

where integer $d \geq 1$ is a specified critical number of paths. However, as illustrated by Figs. 6.22 and 6.23, $Q(r)$ depends on $P^\star$; we cannot know the value of $P^\star$ since it is the true path taken by the vehicle.

We assume $P^\star$ is a random variable that takes values in $\{P_j\}$, $j = 1, 2, \ldots, n$. Under this assumption, Eq. (6.43) can be written as a probability statement, *i.e.*,

$$\text{choose sensor layout and } r \in [r_1, r_2] \text{ such that } P(Q(r) \leq d) = \bar{q}, \tag{6.44}$$

where $\bar{q} \in [0, 1]$ is a specified level of reliability. Note that $Q(r)$ is a stochastic process since, for any fixed r, Q is a random variable. The probability law for $P^\star$ is unknown; four probability laws for $P^\star$ are considered in the following section. Equation (6.44) defines the performance requirement for the sensor network design; the parameters used for calculations are listed in Table 6.7.

Table 6.7. Parameters used in vehicle monitoring problem.

Description	Parameter	Value
Number of possible paths	n	12
Critical number of paths	d	3
Reliability	$\bar{q}$	0.6
Sensor range lower bound	r_1	1
Sensor range upper bound	r_2	15

6.2.4 Sensitivity of model output

The properties of Q, the number of posterior paths in Ω, are sensitive to the properties of $P^\star$, the actual path of the vehicle. To illustrate, consider four possible probability laws for $P^\star$:

$$\begin{aligned}
m_1\colon & P(P^\star = P_k) = 1/12,\ k = 1, 2, \ldots, 12, \\
m_2\colon & P(P^\star = P_k) = \begin{cases} 1/8, & k = 1, 2, 3, 4, \\ 1/16, & k = 5, 6, \ldots, 12, \end{cases} \\
m_3\colon & P(P^\star = P_k) = \begin{cases} 1/6, & k = 1, 5, 9, \\ 1/18, & k = 2, 3, 4, 6, 7, 8, 10, 11, 12, \text{ and} \end{cases} \\
m_4\colon & P(P^\star = P_k) = \begin{cases} 1/6, & k = 2, 6, 10, \\ 1/18, & k = 1, 3, 4, 5, 7, 8, 9, 11, 12. \end{cases}
\end{aligned} \tag{6.45}$$

By Eq. (6.45) and Fig. 6.16, law m_1 assumes each path is equally likely, law m_2 assumes shorter paths are more likely than longer paths, and laws m_3 and m_4 assume paths originating from the south and east, respectively, are more likely than paths originating from the north or west. It follows that the mean, variance, and marginal CDF of Q are given by

$$\begin{aligned}
\mathrm{E}[Q(r)] &= \sum_{j=1}^{n} Q(r|P^\star = P_j) P(P^\star = P_j), \\
\mathrm{Var}[Q(r)] &= \sum_{j=1}^{n} \left[Q(r|P^\star = P_j) - \mathrm{E}[Q(r)] \right]^2 P(P^\star = P_j), \text{ and} \\
P(Q(r) = k) &= \sum_{j=1}^{n} 1\left[Q(r|P^\star = P_j) = k \right] P(P^\star = P_j),\ k = 1, 2, \ldots, n.
\end{aligned} \tag{6.46}$$

These properties are calculated assuming laws m_1, m_2, m_3, and m_4 for layouts #1 and #2; results are shown in Figs. 6.24 and 6.25, respectively. Note that in each case, as r increases, $\mathrm{E}[Q(r)] \to n$ and $\mathrm{Var}[Q(r)] \to 0$. Also, with the exception of layout #1 and $r = 5$, $P(Q(r) \leq k) < 1$ for $k < n$.

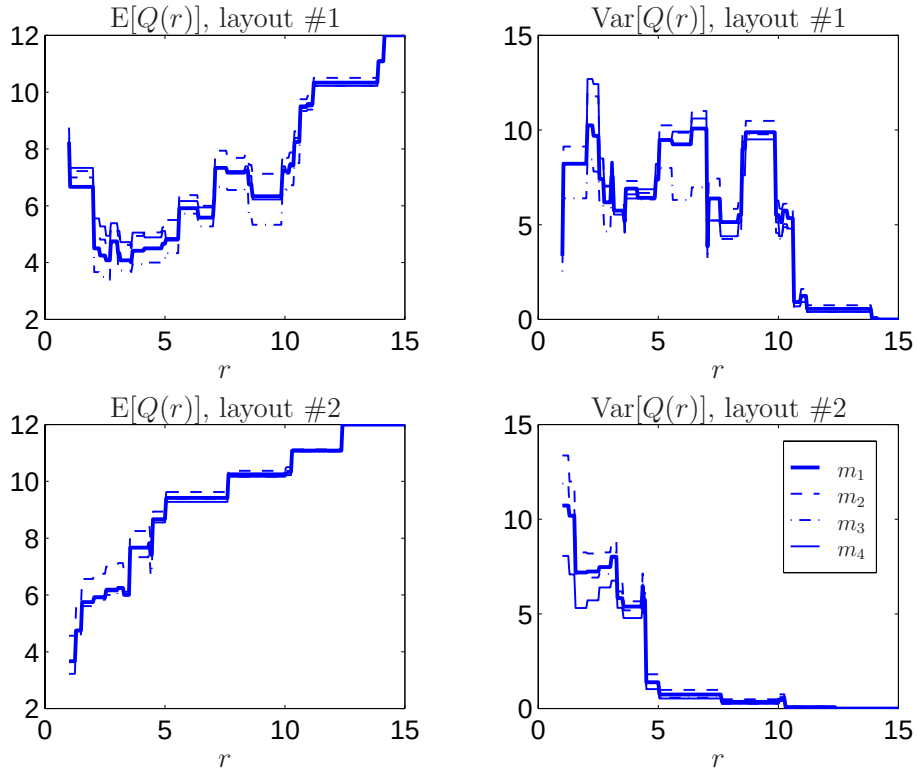

Figure 6.24: Second moment properties of $Q(r)$.

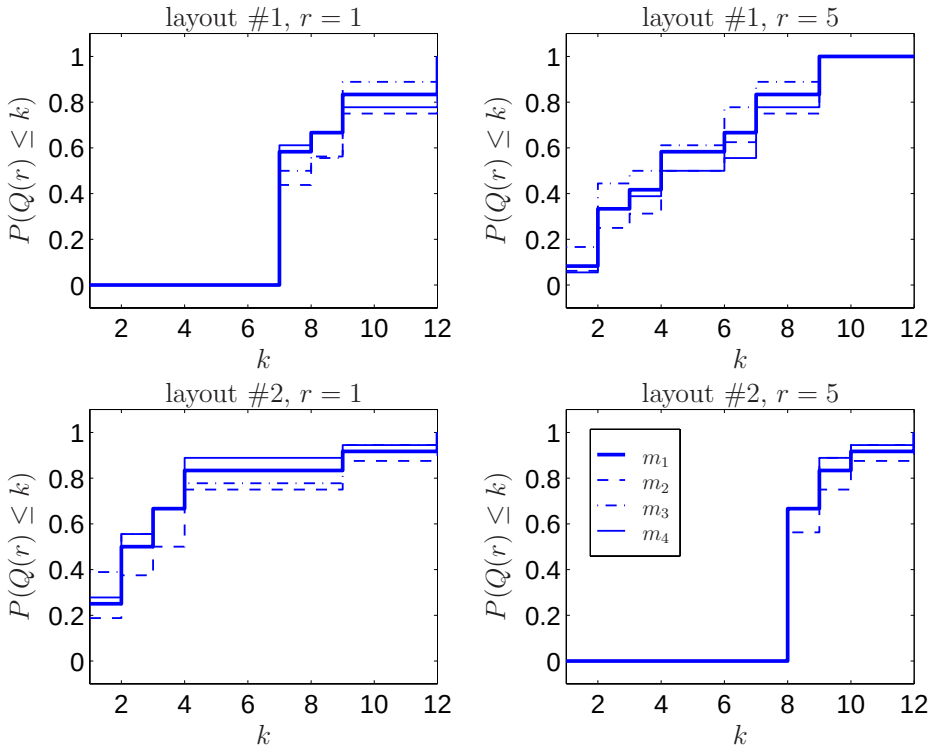

Figure 6.25: Marginal CDFs of $Q(r = 1)$ and $Q(r = 5)$.

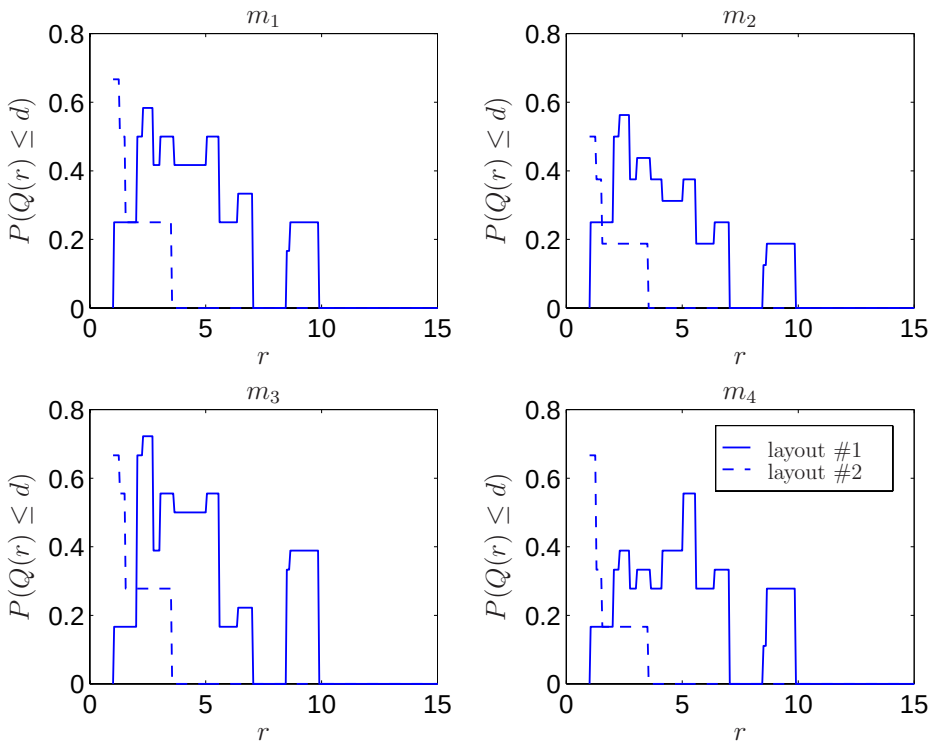

Figure 6.26: Performance of designs for $d = 3$.

The performance of layouts #1 and #2, as defined by Eq. (6.44), is shown in Fig. 6.26 under each m_i, $i = 1, 2, 3, 4$. Under laws m_1 and m_4, layout #2, with $r = 1$, outperforms layout #1 for any r. Likewise, under laws m_2 and m_3, layout #1 with $r \approx 2.5$ outperforms layout #2 for any r. In general, the range, r, needed to satisfy Eq. (6.44) is sensitive to both sensor layout and the probability law of $P^\star$.

6.2.5 Model selection

In this section, we use the decision-theoretic method for model selection to select the optimal design of the sensor network, *i.e.*, optimal sensor layout and range. The model use is design, so the formulation of Section 2.3.2.2 applies. The candidate designs of the sensor network are presented in Section 6.2.5.1; the optimal design is selected in Section 6.2.5.2.

6.2.5.1 Candidate designs

The possible states of nature are the possible probability laws for $P^\star$, *i.e.*,

$$\mathcal{M} = \{m_1, m_2, m_3, m_4\}, \tag{6.47}$$

where m_i, $i = 1, 2, 3, 4$, is defined by Eq. (6.45). We assume noninformative prior information, meaning that $p_i = 1/4$, $i = 1, 2, 3, 4$, where p_i denotes the probability that m_i is true. The choice of candidate probability laws for $P^\star$ in Eq. (6.45) is arbitrary in this example; additional models can be considered without altering the procedure that follows. However, there may be a rational means to assign these laws using, for example, the populations of the cities that lie to the south and east of Ithaca.

The space of possible actions, $\mathcal{A}$, is the space of possible designs of the sensor network. Each design, a_i, includes one of the four sensor layouts shown in Figs. 6.17-6.20, as well as a range, r, for each sensor. Define

$$\begin{aligned}
\mathcal{A}_1 &= \left\{ \text{layout } \#1,\, r_i = \frac{i-1}{20} + 1, i = 1, 2, \ldots, 281 \right\}, \\
\mathcal{A}_2 &= \left\{ \text{layout } \#2,\, r_i = \frac{i-1}{20} + 1, i = 1, 2, \ldots, 281 \right\}, \\
\mathcal{A}_3 &= \left\{ \text{layout } \#3,\, r_i = \frac{i-1}{20} + 1, i = 1, 2, \ldots, 281 \right\}, \\
\mathcal{A}_4 &= \left\{ \text{layout } \#4,\, r_i = \frac{i-1}{20} + 1, i = 1, 2, \ldots, 281 \right\},
\end{aligned} \tag{6.48}$$

where each $\mathcal{A}_i$, $i = 1, 2, 3, 4$, contains 281 elements due to the discretization of $r \in [r_1, r_2]$. Three cases are then considered; **case (i)**: $\mathcal{A} = \{\mathcal{A}_1, \mathcal{A}_2\}$, **case (ii)**: $\mathcal{A} = \{\mathcal{A}_1, \mathcal{A}_3\}$, and **case (iii)**: $\mathcal{A} = \{\mathcal{A}_2, \mathcal{A}_4\}$. For case (i), we assume the number of sensors is fixed at $q = 9$ due to cost constraints; methods of model selection can be used to determine the optimal sensor layout and range. For case (ii), we assume the sensors cannot be placed on any edge in the graph of Ω, *i.e.*, any of the roads in New York state; a rectangular arrangement of sensors is considered and methods of model selection are used to select the optimal number of sensors, q, and optimal sensor range, r. Likewise, for case (iii), we assume the sensors must be placed on the edges of the graph of Ω and use model selection techniques to determine the optimal q and r.

6.2.5.2 Optimal design

To select the optimal design for the sensor network, we use the formulation of Section 2.3.2.2. The utility function, defined for each design, $a_i \in \mathcal{A}$, and each state of nature, $m_j \in \mathcal{M}$, is given by

$$U(a_i, m_j) = \gamma(a_i) + \psi(a_i, m_j), \tag{6.49}$$

where $\gamma(a_i) \geq 0$ denotes the cost of design a_i, and $\psi(a_i, m_j) \geq 0$ denotes the penalty associated with design a_i, if m_j is true.

Table 6.8. Optimal sensor network design for cases (i), (ii), and (iii).

Case	Optimal design, $a^\star$		
	Number of sensors	Sensor layout	Sensor range
(i)	$q = 9$	#2	$r = 1$
(ii)	$q = 12$	#3	$r = 5.15$
(iii)	$q = 9$	#2	$r = 1$

Let q_i and r_i be the number of sensors and sensor range, respectively, under design a_i. The cost function is given by

$$\gamma(a_i) = \tilde{\gamma}(q_i) = \gamma_0 \; q_i, \tag{6.50}$$

where $\gamma_0 \geq 0$ is a deterministic parameter. By Eq. (6.50), designs with a large number of sensors (layouts #3 and #4) have a high cost. Consider the case of design a_i and model m_j. We define ξ_{ij} such that

$$P(Q(r_i) \leq \xi_{ij} \mid m_j) = \bar{q}. \tag{6.51}$$

The penalty function is given by

$$\psi(a_i, m_j) = \tilde{\psi}(\xi_{ij}, d) = \begin{cases} \beta_1(\xi_{ij} - d)^2 & \text{if } \xi_{ij} \leq d, \text{ and} \\ \beta_2(\xi_{ij} - d)^2 & \text{if } \xi_{ij} > d, \end{cases} \tag{6.52}$$

where $\beta_2 \geq \beta_1 > 0$ are deterministic parameters; Eq. (6.52) is identical to Eq. (2.53) in Section 2.3.2.2. Parameters $\gamma_0 = 1/10$, $\beta_1 = 0$, and $\beta_2 = 1$ are used for calculations.

The expected utility of design a_i is given by Eq. (2.45), *i.e.*,

$$u(a_i) = \gamma(a_i) + \sum_{j=1}^{4} \psi(a_i, m_j) \; p_j, \;\; i = 1, 2, \ldots, 562. \tag{6.53}$$

The expected utility of each design was calculated for each of the three cases considered; results are shown in Fig. 6.27. For each case, the optimal design, $a^\star$, minimizes Eq. (6.53); results are listed in Table 6.8. Note that the values for $u(a_i)$, $i = 1, 2, \ldots, 281$, coincide for cases (i) and (ii) because $\mathcal{A} = \{\mathcal{A}_1, \mathcal{A}_2\}$ and $\mathcal{A} = \{\mathcal{A}_1, \mathcal{A}_3\}$ for cases (i) and (ii), respectively.

When possible, it was advantageous to place sensors on the edges of the graph of Ω rather than in a rectangular grid. In this case, the smallest possible value for r was optimal, and additional sensors did not significantly improve the design. For the case when the sensors were restricted from being on any edge (case (ii)), it was advantageous to add more sensors to the network, and use a larger range.

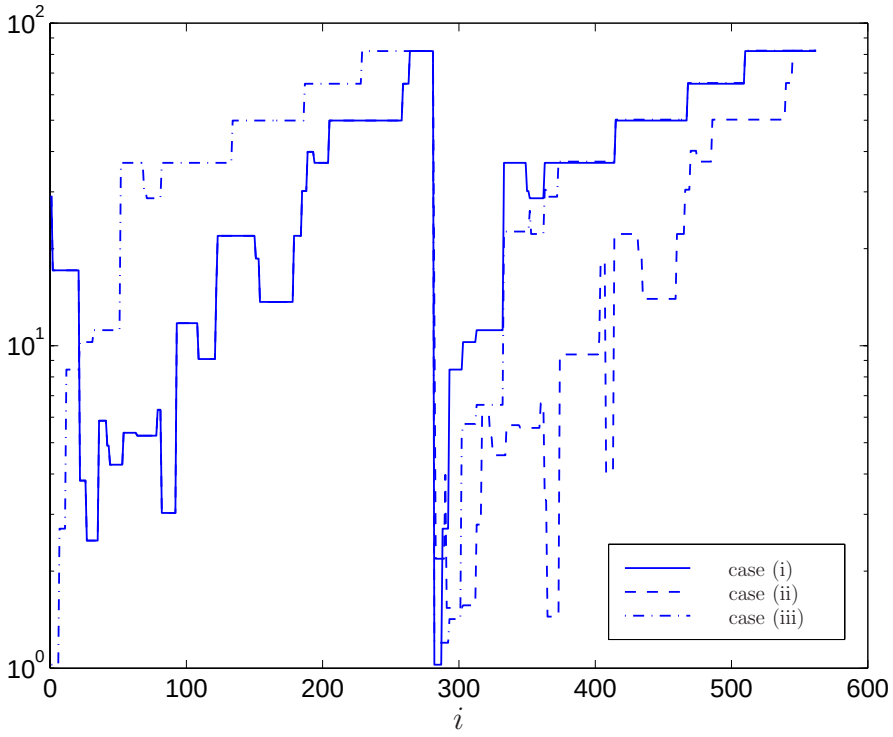

Figure 6.27: Expected utilities, $u(a_i)$, for cases (i), (ii), and (iii).

6.2.6 Future work

Possible extensions to the vehicle monitoring problem are discussed. In Section 6.2.1.1, we assumed there was a single treasure in Ω, the city of Ithaca. Further, we assumed the vehicle paths to be independent of the sensor locations in Section 6.2.1.2. These assumptions were made to simplify the calculations; if they are relaxed or removed, the models for the possible paths and vehicle movement within Ω will be more complicated. The benefits are numerous; we may construct a mathematical model to describe, for example, the movement of vehicles that have some device to detect the presence of a sensor on their intended route of travel, *e.g.*, radar detectors.

By Eqs. (6.38) and (6.39), it is clear that no information on the time a vehicle passes the sensor is available; the sensor is either activated or it is not. If time is included in the model, we may achieve: (1) knowledge of vehicle location at any time, and (2) improved monitoring performance. To illustrate the latter, consider Fig. 6.28, which shows three possible paths through Ω, labeled P_1, P_2, and P_3. Three sensors, labeled A, B, and C, are positioned to monitor vehicle traffic in Ω; the ranges of sensors A, B, and C intersect. P_1, P_2, and P_3 pass through the range of each sensor and will therefore activate all three sensors. The time-invariant sensor model developed in Section 6.2.1.3 is unable to distinguish among the three possible paths because the path metric defined by Eq. (6.39) is identical for paths P_1, P_2, and P_3. However, we note that P_1 is the only path that activates sensors A and B simultaneously, P_2 is the only path that activates sensors B and C simultaneously,

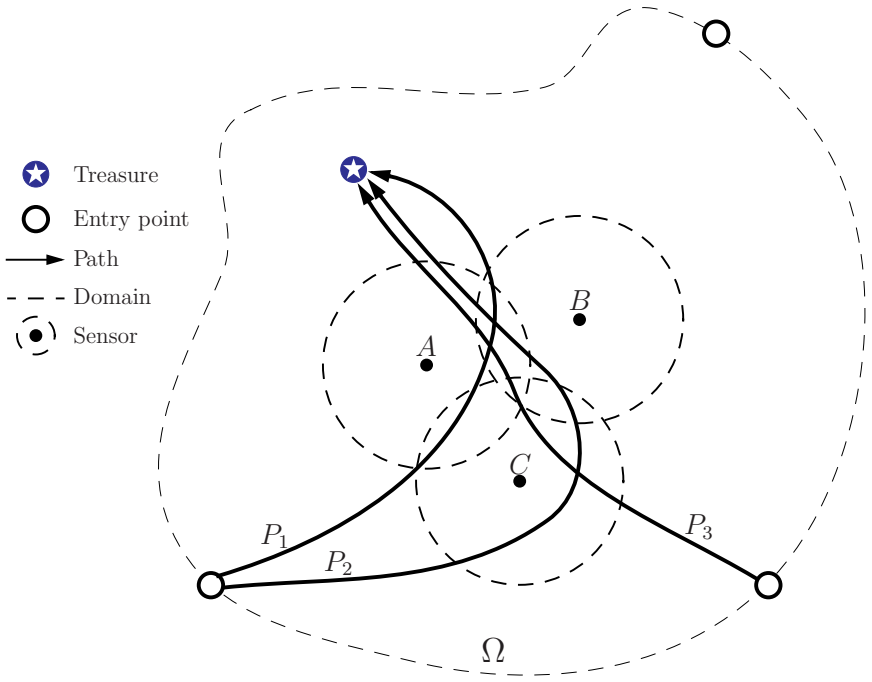

Figure 6.28: Scenario to illustrate time-varying sensor model.

and P_3 is the only path that activates all sensors simultaneously. A time-varying model for the sensor can utilize this information; improvements in vehicle monitoring will result.

This page intentionally left blank.

Chapter 7

Conclusions

Mathematical models are developed and used to study the properties of complex systems and/or modify these systems to satisfy some performance requirements in just about every area of applied science and engineering. A particular reason for which a model is developed, *e.g.*, performance assessment or design, was referred to as the model use. As discussed in Chapter 1, information on the system being modeled is, in general, incomplete, so that there may be two or more models consistent with the available information. The collection of these models was called the class of candidate models. Our objective was the development of a methodology for selecting the optimal member from a class of candidate models for the system that is sufficiently accurate for an intended use.

In Chapter 2, generalities and essentials of the model selection problem were presented. Three methods for analysis were discussed: (1) the method of maximum likelihood, (2) Bayesian methods, and (3) a decision-theoretic method. Classical techniques for model selection include methods (1) and (2); these techniques ignore model use and require data to be available. Method (3) requires elements from decision theory, most notably the concept of a utility function, which allowed the consequences of using an inappropriate model for the system to be quantified. General utility functions, for the case of model use for design and performance prediction, were developed. It was shown that the optimal model depended on the: (i) available information, (ii) collection of candidate models, and in the case of the decision-theoretic method for model selection, (iii) model use. Further, the methods for model selection considered did not require that the candidate models be random.

Methods for model selection were applied to simple, relevant example problems in Chapter 3. It was shown that, for the examples considered, classical methods for model selection were unreliable when data was limited. The decision-theoretic approach to model selection did not have this limitation, and model use was included in the selection process through an appropriate utility function. This feature was especially important when modeling high risk systems, where the consequences of

using an inappropriate model could be disastrous.

One class of candidate model, the polynomial chaos (PC) approximation, was defined and studied in Chapter 4 for the case of non-Gaussian random variables and stationary stochastic processes. It was shown that: (1) the accuracy of the PC approximation depended on the number of terms retained in the series, (2) the accuracy did not necessarily improve when more terms were retained, and (3) the L_2 convergence properties of the PC approximation were largely irrelevant for applications since, generally, the PC approximation was defined by a small number of terms. The decision-theoretic method for model selection was developed to select the optimal number of terms for the PC approximation. For illustration, a collection of PC approximations for a non-Gaussian random variable was considered. A utility function was used to assign a penalty to those PC approximations that were inaccurate and/or required a large number of terms.

The decision-theoretic method for model selection was applied to select optimal pressure load models acting on a spacecraft during atmospheric re-entry in Chapter 5. This is one example of a high risk system. The collection of candidate models formed a class of stationary stochastic processes in space and time with partially known probability laws. Utility functions were formulated based on properties of the response of an internal component. Three cases were considered: (1) each candidate model was assumed Gaussian with perfectly known temporal correlation function and spatial correlation function of specified functional form, but with an unknown parameter value, (2) each candidate model was assumed Gaussian with perfectly known temporal correlation function and spatial correlation function of unspecified functional form with unknown parameters, and (3) each candidate model was assumed to be completely defined in the second-moment sense, but with unknown marginal PDF. It was shown that the internal component response is sensitive to the degree of spatial correlation on the input for cases (1) and (2), and to the marginal PDF of the input for case (3). Further, the Gaussian model that is uncorrelated in the spatial dimension, a model that is often applied for this problem, is not optimal.

In Chapter 6, another high risk system, namely the design of a distributed sensor network, was considered. The detection and classification of vehicles by a single sensor was addressed in Section 6.1, while the monitoring of vehicles by a collection of sensors within a two-dimensional domain was address in Section 6.2. In both cases, models for vehicle movement and the sensor network, as well as metrics of performance, were developed; the information on vehicle movement was incomplete. The decision-theoretic method for model selection was then applied to select the optimal design of the sensor network. A collection of candidate designs for the sensor network was defined, and postulated utility functions were used to penalize any candidate design that failed to achieve the required levels of performance.

References

[1] B. Abraham and J. Ledolter. *Statistical Methods for Forecasting*. Wiley, 1983.

[2] M. Abramowitz and I.A. Stegun. *Handbook of Mathematical Functions*. Dover Publications, Inc., New York, NY, 1972.

[3] K.F. Alvin, W.L. Oberkampf, B.M. Rutherford, and K.V. Diegert. Methodology for characterizing modeling and discretization uncertainties in computational simulation. Technical Report SAND2000-0515, Sandia National Laboratories, 2000.

[4] American Institute of Aeronautics and Astronautics. *AIAA Guide for the Verification and Validation of Computational Fluid Dynamics Simulations*. Number AIAA-G-077-1998. American Institute of Aeronautics & Astronautics, 1998.

[5] A. Ang and W. Tang. *Probability Concepts in Engineering Planning and Design: Vol. 1 - Basic Principles*. John Wiley & Sons, Inc., New York, NY, 1975.

[6] T. Banek. Chaos expansion for the solution of stochastic differential equations. *Systems & Control Letters*, 36:351–358, 1999.

[7] J.L. Beck and L.S. Katafygiotis. Updating models and their uncertainties I: Bayesian statistical framework. *Journal of Engineering Mechanics*, 124(4):455–461, 1998.

[8] J.L. Beck and K.-V. Yuen. Model selection using response measurements: Bayesian probabilistic approach. *Journal of Engineering Mechanics*, 130(2):192–203, 2004.

[9] G. Becker and A. Güdesen. Passive sensing with acoustics on the battlefield. *Applied Acoustics*, 59:149–178, 1999.

[10] L.A. Bergman and J.W. Nicholson. Stationary response of combined linear dynamic systems to stationary excitation. *Journal of Sound and Vibration*, 103(2):225–236, 1985.

[11] J.M. Bernardo and A.F.M. Smith. *Bayesian Theory*. Wiley, 1994.

[12] S. Blackman and R. Popoli. *Design and Analysis of Modern Tracking Systems*. Artech House, Boston, 1999.

[13] D. Blackwell and M.A. Girshick. *Theory of Games and Statistical Decisions*. Dover, 1954.

[14] R.C. Blanchard and J.Y. Nicholson. Orbiter rarefied-flow reentry measurements from the OARE on STS-62. Technical Report TM-110182, NASA, 1995.

[15] G.E.P. Box and D.W. Behnken. Some new three level designs for the study of quantitative variables. *Technometrics*, 2(4):455–475, 1960.

[16] P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods*. Springer, New York, NY, 1991.

[17] M.K. Bull. Wall-pressure fluctuations beneath turbulent boundary layers: Some reflections on forty years of research. *Journal of Sound and Vibration*, 190(3):299–315, 1996.

[18] K.P. Burnham and D.R. Anderson. *Model Selection and Inference: A Practical Information-Theoretic Approach*. Springer, 1998.

[19] E.M. Carapezza, editor. *Unattended Ground Sensor Technologies and Applications III*, volume 4393, Orlando, FL, April 18-19 2001. SPIE - The International Society for Optical Engineering.

[20] E.M. Carapezza, D.B. Law, and K. Terry Stalker, editors. *Unattended Ground Sensor Technologies and Applications*, volume 3713, Orlando, FL, April 8-9 1999. SPIE - The International Society for Optical Engineering.

[21] H. Chernoff and L.E. Moses. *Elementary Decision Theory*. Dover Publications, Inc., New York, NY, 1959.

[22] P.F. Christoffersen and F.X. Diebold. Further results on forecasting and model selection under asymmetric loss. *Journal of Applied Econometrics*, 11:561–571, 1996.

[23] R.R. Clements. *Mathematical Modelling: A Case Study Approach*. Cambridge University Press, 1989.

[24] J.A. Cockburn and J.E. Robertson. Vibration response of spacecraft shrouds to in-flight fluctuating pressures. *Journal of Sound and Vibration*, 33(4):399–425, 1974.

[25] R.R. Craig, Jr. *Structural Dynamics*. John Wiley & Sons, Inc., New York, NY, 1981.

[26] N.A.C. Cressie. *Statistics for Spatial Data*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, NY, 1993.

[27] O. Ditlevsen and H.O. Madsen. *Structural Reliability Methods*. O. Ditlevsen and H.O. Madsen, http://www.mek.dtu.dk/staff/od/books.htm, internet 2.1 edition, 2002.

[28] I. Elishakoff. *Probabilistic Theory of Structures*. Dover, 1999.

[29] I. Elishakoff, A. Th. van Zanten, and S. H. Crandall. Wide-band random axisymmetric vibration of cylindrical shells. *Journal of Applied Mechanics*, 46:417–422, 1979.

[30] L.D. Ely. *Return from Space: An Explanation of Re-entry Problems and Factors of Re-entry Vehicle Design and Performance*. Charles C. Thomas, publisher, Springfield, IL, 1966.

[31] W. Feller. *An Introduction to Probability Theory and Its Applications, Vol. I*. John Wiley & Sons, Inc., New York, NY, third edition, 1968.

[32] R.V. Field, Jr. Numerical methods to estimate the coefficients of the polynomial chaos expansion. In *2002 Engineering Mechanics Conference*, New York, NY, 2002. American Society of Civil Engineers (ASCE).

[33] R.V. Field, Jr. and M. Grigoriu. A new perspective on polynomial chaos. In P.D. Spanos and G. Deodatis, editors, *Fourth International Computational Stochastic Mechanics Conference*, pages 199–205, Corfu, Greece, June 9-12 2002. Millpress.

[34] R.V. Field, Jr., M. Grigoriu, and S.F. Wojtkiewicz. Probabilistic model selection for systems with design constraints. In A. Der Kiureghian, S. Madanat, and J. Pestana, editors, *Applications of Statistics and Probability in Civil Engineering*, pages 435–442, San Francisco, CA, July 6-9 2003. Millpress.

[35] R.A. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London, Series A*, 222:309–368, 1922.

[36] W. Gautschi. *Numerical Analysis: An Introduction*. Birkhäuser, Boston, MA, 1997.

[37] N. Gershenfeld. *The Nature of Mathematical Modeling*. Cambridge University Press, 1999.

[38] G. Geyskens and A. Der Kiureghian. Bayesian assessment and selection of models for structural reliability analysis. In *Probabilistic Mechanics and Structural and Geotechnical Reliability*, pages 566–569, 1996.

[39] R.G. Ghanem. The nonlinear Gaussian spectrum of log-normal stochastic processes and variables. *Journal of Applied Mechanics*, 66:964–973, 1999.

[40] R.G. Ghanem and S. Dham. Stochastic finite element analysis for multiphase flow in heterogeneous porous media. *Transport in Porous Media*, 32:239–262, 1998.

[41] R.G. Ghanem and J.R. Red-Horse. Propagation of probabilistic uncertainty in complex physical systems using a stochastic finite element approach. *Physica D*, 133:137–144, 1999.

[42] R.G. Ghanem and P. D. Spanos. Stochastic Galerkin expansion for nonlinear random vibration analysis. *Probabilistic Engineering Mechanics*, 8(3):255–264, 1993.

[43] R.G. Ghanem and P.D. Spanos. *Stochastic Finite Elements: A Spectral Approach*. Springer-Verlag, New York, NY, 1991.

[44] A.A. Giunta and L.T. Watson. A comparison of approximation modeling techniques: Polynomial versus interpolating models. *AIAA Journal*, 1998.

[45] G. L Goodman. Detection and classification for unattended ground sensors. In Robin Evans, Lang White, Daniel McMichael, and Len Sciacca, editors, *Proceedings of Information Decision and Control 99*, pages 419–424, Adelaide, Australia, February 1999. Institute of Electrical and Electronic Engineers, Inc.

[46] M. Grigoriu. *A Decision Theoretic Approach to Model Selection for Structural Reliability*. PhD thesis, Massachusetts Institute of Technology, 1976.

[47] M. Grigoriu. Crossings of non-Gaussian translation processes. *Journal of Engineering Mechanics*, 110:610–620, 1984.

[48] M. Grigoriu. *Applied Non-Gaussian Processes*. P T R Prentice-Hall, Englewood Cliffs, NJ, 1995.

[49] M. Grigoriu. A spectral representation based model for Monte Carlo simulation. *Probabilistic Engineering Mechanics*, 15:365–370, 2000.

[50] M. Grigoriu. *Stochastic Calculus: Applications in Science and Engineering*. Birkhäuser, Boston, MA, 2002.

[51] M. Grigoriu, D. Veneziano, and C.A. Cornell. Probabilistic modeling as decision making. *Journal of Engineering Mechanics*, 105(4):585–596, 1979.

[52] D. L. Hall. *Mathematical Techniques in Multisensor Data Fusion*. Artech House, Boston, 1992.

[53] B. Hassan. Estimation of fluctuation pressures on a sphere-cone geometry at hypersonic speeds. Technical report, Sandia National Laboratories, September 2001.

[54] G.A. Hazelrigg. *Systems Engineering: An Approach to Information-Based Design*. Prentice Hall, 1996.

[55] M. Hetényi. *Beams on Elastic Foundation*. The University of Michigan Press, 1983.

[56] K.D. Hjelmstad. *Fundamentals of Structural Mechanics*. Prentice Hall, 1997.

[57] J.S.U. Hjorth. *Computer Intensive Statistical Methods: Validation, Model Selection, and Bootstrap*. Chapman & Hall, 1994.

[58] T. Igusa, S. Buonopane, and B. Ellingwood. Bayesian analysis of uncertainty for structural engineering applications. *Structural Safety*, 24:165–186, 2002.

[59] C. Kafali and M. Grigoriu. Non-Gaussian model for spatially coherent seismic ground motions. In A. Der Kiureghian, S. Madanat, and J. Pestana, editors, *Applications of Statistics and Probability in Civil Engineering*, pages 321–327, San Francisco, CA, July 6-9 2003. Millpress.

[60] G. Kallianpur. *Stochastic Filtering Theory*. Springer-Verlag, New York, NY, 1980.

[61] J.N. Kapur. *Mathematical Modelling*. John Wiiley & Sons, Inc., 1988.

[62] L.S. Katafygiotis and J.L. Beck. Updating models and their uncertainties II: Model indentifiability. *Journal of Engineering Mechanics*, 124(4):463–467, 1998.

[63] E.S. Keeping. *Introduction to Statistical Inference*. Dover, 1995.

[64] A.I. Khuri and J.A. Cornell. *Response Surfaces: Designs and Analyses*. Marcel Dekker, Inc., 1996.

[65] H.T. Kung and D. Vlah. Efficient location tracking using sensor networks. In *Proceedings of 2003 IEEE Wireless Communications and Networking Conference*, 2003.

[66] A.L. Laganelli, A. Martellucci, and L.L Shaw. Wall pressure fluctuations in attached boundary-layer flow. *AIAA Journal*, 21(4):495–502, 1983.

[67] K.B. Laskey. Model uncertainty: Theory and practical implications. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 26(3):340–348, 1996.

[68] L. Ljung and T. Glad. *Modeling of Dynamic Systems*. Prentice Hall, Inc., 1994.

[69] D. Maki and M. Thompson. *The Mathematical Modeling Cycle*. http://www.indiana.edu/˜hmathmod/modelmodel.html.

[70] K.J. Martijn Cremers. Stock return predictability: A Bayesian model selection perspective. *The Review of Financial Studies*, 15(4):1223–1249, 2002.

[71] J.J. Martin. *Atmospheric Reentry: An Introduction to its Science and Engineering*. Prentice Hall, Inc., Englewood Cliffs, NJ, 1966.

[72] H.G. Matthies, C.E. Brenner, C.G. Bucher, and C.G. Soares. Uncertainties in probabilistic numerical analysis of structures and solids: Stochastic finite elements. *Structural Safety*, 19(3):283–336, 1997.

[73] Russell Merris. *Graph Theory*. John Wiley & Sons, New York, 2001.

[74] H. Mitchell and M.D. McKenzie. GARCH model selection criteria. *Quantitative Finance*, 3:262–284, 2003.

[75] W.L. Oberkampf, T.G. Trucano, and C. Hirsch. Verification, validation, and predictive capability in computational engineering and physics. Technical Report SAND2003-3769, Sandia National Laboratories, 2003.

[76] C. Papadimitriou, J. Beck, and L. Katafygiotis. Updating robust reliability using structural test data. *Probabilistic Engineering Mechanics*, 16:103–113, 2001.

[77] K.K. Phoon, S.P. Huang, and S.T. Quek. Implementation of Karhunen-Loève expansion for simulation using a wavelet-Galerkin scheme. *Probabilistic Engineering Mechanics*, 17:293–303, 2002.

[78] B. Puig, F. Poirion, and C. Soize. Non-Gaussian simulation using Hermite polynomial expansion: Convergence and algorithms. *Probabilistic Engineering Mechanics*, 17:253–264, 2002.

[79] H. Qi, S. Iyengar, and K. Chakrabarty. Distributed sensor networks - a review of recent research. *Journal of the Franklin Institute*, 338:655–668, 2001.

[80] A.E. Raftery, D. Madigan, and J.A. Hoeting. Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, 92:179–191, 1997.

[81] H. Raiffa and R. Schlaifer. *Applied Statistical Decision Theory*. Division of Research, Harvard Business School, 1961.

[82] J.R. Red-Horse and R.G. Ghanem. Polynomial chaos representation of the random eigenvalue problem. In *AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*. American Institute of Aeronautics and Astronautics (AIAA), 1999.

[83] S.I. Resnick. *Adventures in Stochastic Processes*. Birkhäuser, Boston, 1992.

[84] C.P. Robert. *The Bayesian Choice*. Springer, 2001.

[85] S. Ross. *A First Course in Probability*. Macmillan College Publishing Company, New York, NY, 1994.

[86] S. Sakamoto and R.G. Ghanem. Polynomial chaos decomposition for the simulation of non-Gaussian non-stationary stochastic processes. *Journal of Engineering Mechanics*, 128(2):190–201, 2002.

[87] S. Schlesinger. Terminology for model credibility. *Simulation*, 32(3):103–104, 1979.

[88] T.T. Soong and M. Grigoriu. *Random Vibration of Mechanical and Structural Systems*. P T R Prentice-Hall, Englewood Cliffs, NJ, 1993.

[89] T.G. Trucano, R.G. Easterling, K.J. Dowding, T.L. Paez, A. Urbina, V.J. Romero, B.M. Rutherford, and R.G. Hills. Description of the Sandia validation metrics project. Technical Report SAND2001-1339, Sandia National Laboratories, 2001.

[90] J.M. van Noortwijk, H.J. Kalk, and M.T. Duits. The use of Bayes factors for model selection in structural reliability. In *8th International Conference on Structural Safety and Reliability (ICOSSAR)*, June 17-21 2001.

[91] H.L. Van Trees. *Detection, Estimation, and Modulation Theory, Part 1*. Wiley, New York, NY, 1968.

[92] A. Vehtari. *Bayesian Model Assessment and Selection Using Expected Utilities*. PhD thesis, Helsinki University of Technology, 2001.

[93] J. von Neumann and O. Morgenstern. *Theory of Games and Economical Behavior*. Princeton University Press, Princeton, 1944.

[94] N. Wiener. The homogeneous chaos. *American Journal of Mathematics*, 60(4):897–936, 1938.

[95] S.R. Winterstein. Non-normal responses and fatigue damage. *Journal of Engineering Mechanics*, 111(10):1291–1295, 1985.

[96] G. Yonas, editor. *Peace and Wartime Applications and Technical Issues for Unattended Ground Sensors*, volume 3081, Orlando, FL, April 22-23 1997. SPIE - The International Society for Optical Engineering.

[97] K.-V. Yuen and J. Beck. Updating properties of nonlinear dynamical systems with uncertain input. *Journal of Engineering Mechanics*, 129(1):9–20, 2003.

[98] A. Zellner. *An Introduction to Bayesian Inference in Econometrics*. John Wiley & Sons, Inc., 1971.

[99] A. Zerva and V. Zervas. Spatial variation of seismic ground motions: An overview. *Applied Mechanics Review*, 55(3):271–297, 2002.

[100] J. Zhang and B. Ellingwood. Orthogonal series expansions of random fields in reliability analysis. *Journal of Engineering Mechanics*, 120(12):2660–2677, 1994.

[101] R. Zhang and S. Mahadevan. Model uncertainty and Bayesian updating in reliability-based inspection. *Structural Safety*, 22:145–160, 2000.

This page intentionally left blank.

Appendix A

Derivation of polynomial chaos coefficients

Consider the reflected Gaussian random variable of Table 4.2, given by $g_3(W) = |W|$. Let

$$|W| = \sum_{i=0}^{\infty} y_i h_i(W) \tag{A.1}$$

be the PC representation of this random variable in terms of the Hermite polynomials, $h_i(W)$. The even and odd Hermite polynomials can be written explicitly [2]

$$\begin{aligned} h_{2k}(s) &= \sum_{j=0}^{k} a_j s^{2(k-j)} \\ h_{2k+1}(s) &= \sum_{j=0}^{k} b_j s^{2(k-j)+1}, \ k = 0, 1, 2, \ldots, \end{aligned} \tag{A.2}$$

where $a_0 = b_0 = 1$ and, for $j = 1, 2, \ldots, k$,

$$\begin{aligned} a_j &= (-1)^j \frac{(2k)!}{j! 2^j (2k-2j)!} \\ b_j &= (-1)^j \frac{(2k+1)!}{j! 2^j (2k-2j+1)!} \end{aligned} \tag{A.3}$$

The coefficients, $\{y_i\}$, of the PC approximation for Eq. (A.1) are given by Eq. (4.8), *i.e.*,

$$\begin{aligned} y_i &= \frac{\mathrm{E}\left[|W| h_i(W)\right]}{\mathrm{E}\left[h_i(W)^2\right]} = \frac{1}{i!} \mathrm{E}\left[|W| h_i(W)\right] \\ &= \frac{1}{i! \sqrt{2\pi}} \int_{-\infty}^{\infty} |s| h_i(s) e^{-s^2/2} \mathrm{d}s, \ i = 0, 1, 2, \ldots, p. \end{aligned} \tag{A.4}$$

Hence, the even PC coefficients are given by

$$\begin{aligned} y_{2k} &= \frac{1}{\sqrt{2\pi}(2k)!}\int_{-\infty}^{\infty}|s|h_{2k}(s)e^{-s^2/2}\mathrm{d}s \\ &= \frac{1}{\sqrt{2\pi}(2k)!}\left\{\sum_{j=0}^{k}a_j\int_{-\infty}^{\infty}|s|s^{2(k-j)}e^{-s^2/2}\mathrm{d}s\right\} \\ &= \frac{2}{\sqrt{2\pi}(2k)!}\left\{\sum_{j=0}^{k}a_j\int_{0}^{\infty}s^{2(k-j)+1}e^{-s^2/2}\mathrm{d}s\right\},\ k=0,1,2,\ldots \end{aligned} \tag{A.5}$$

Let $t = s^2/2$; Eq. (A.5) is rewritten as

$$\begin{aligned} y_{2k} &= \frac{2}{\sqrt{2\pi}(2k)!}\left\{\sum_{j=0}^{k}a_j\int_{0}^{\infty}(2t)^{k-j}e^{-t}\mathrm{d}t\right\} \\ &= \frac{2}{\sqrt{2\pi}(2k)!}\left\{\sum_{j=0}^{k}2^{k-j}a_j\Gamma(k-j+1)\right\} \\ &= \frac{2}{\sqrt{2\pi}}\left\{\frac{2^k k!}{(2k)!}+\sum_{j=1}^{k}(-1)^j\frac{2^{k-2j}(k-j)!}{j!(2k-2j)!}\right\},\ k=0,1,2,\ldots \end{aligned} \tag{A.6}$$

Likewise, the odd PC coefficients are given by

$$\begin{aligned} y_{2k+1} &= \frac{1}{\sqrt{2\pi}(2k+1)!}\int_{-\infty}^{\infty}|s|h_{2k+1}(s)e^{-s^2/2}\mathrm{d}s \\ &= \frac{1}{\sqrt{2\pi}(2k+1)!}\left\{\sum_{j=0}^{k}b_j\int_{-\infty}^{\infty}|s|s^{2(k-j)+1}e^{-s^2/2}\mathrm{d}s\right\} \\ &= \frac{1}{\sqrt{2\pi}(2k+1)!}\left\{\sum_{j=0}^{k}b_j\int_{0}^{\infty}s^{2(k-j)+2}e^{-s^2/2}\mathrm{d}s+\int_{-\infty}^{0}-s^{2(k-j)+2}e^{-s^2/2}\mathrm{d}s\right\} \\ &= 0,\ k=0,1,2,\ldots \end{aligned} \tag{A.7}$$

Appendix B

Simulation of Gaussian vector processes

Let $\mathbf{G}(t)$ be an n-dimensional stationary Gaussian process such that the coordinates of $\mathbf{G}(t)$ have zero mean and covariance functions

$$c_{kl}(\tau) = \mathrm{E}[G_k(t)G_l(t+\tau)],\ k,l = 1,2,\ldots,n. \tag{B.1}$$

The spectral density of $\mathbf{G}(t)$, denoted by $\mathbf{s}(\nu)$, has coordinates ([88], p. 28),

$$s_{kl}(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} c_{kl}(\tau)e^{-i\nu\tau}\mathrm{d}\nu,\ k,l = 1,2,\ldots,n, \tag{B.2}$$

where $i = \sqrt{-1}$. By the spectral representation theorem ([50], pp. 154-155), $\mathbf{G}(t)$ can be expressed as

$$\mathbf{G}(t) = \int_0^{\infty} \left[\cos(\nu t)\mathrm{d}\mathbf{U}(\nu) + \sin(\nu t)\mathrm{d}\mathbf{V}(\nu)\right], \tag{B.3}$$

where $\mathbf{U}$ and $\mathbf{V}$ are zero mean Gaussian processes with orthogonal increments such that, for $k,l = 1,2,\ldots,n$,

$$\begin{aligned}
\mathrm{E}[\mathrm{d}U_k(\nu)\mathrm{d}U_l(\nu')] = &\quad \mathrm{E}[\mathrm{d}V_k(\nu)\mathrm{d}V_l(\nu')] = \delta(\nu-\nu')g_{kl}(\nu)\mathrm{d}(\nu),\\
\mathrm{E}[\mathrm{d}U_k(\nu)\mathrm{d}V_l(\nu')] = &-\mathrm{E}[\mathrm{d}V_k(\nu)\mathrm{d}U_l(\nu')] = \delta(\nu-\nu')h_{kl}(\nu)\mathrm{d}(\nu),
\end{aligned} \tag{B.4}$$

where $\delta(\xi) = 1$ for $\xi = 0$ and zero otherwise, and

$$\begin{aligned}
g_{kl}(\nu) &= s_{kl}(\nu) + s_{kl}(-\nu)\\
h_{kl}(\nu) &= -i\left[s_{kl}(\nu) - s_{kl}(-\nu)\right].
\end{aligned} \tag{B.5}$$

By Eq. (B.3), $\mathbf{G}(t)$ can be viewed as a superposition of harmonics with random (Gaussian) amplitudes.

Define, for each $k = 1, 2, \ldots, n$, a cut-off frequency $\nu_k^\star$, such that

$$\int_{-\nu_k^\star}^{\nu_k^\star} s_{kk}(\nu)\mathrm{d}\nu \approx \int_{-\infty}^{\infty} s_{kk}(\nu)\mathrm{d}\nu, \tag{B.6}$$

and let $\nu^\star = \max_{1\leq k\leq n} \nu_k^\star$. Let (u_{j-1}, u_j), $j = 1, 2, \ldots, q$, with $u_0 = 0$ and $u_q = \nu^\star$, be a partition of $(0, \nu^\star)$ in q nonoverlapping intervals of length $\Delta\nu_j = u_j - u_{j-1}$. Denote by ν_j the midpoint of (u_{j-1}, u_j), $j = 1, 2, \ldots, q$. The following is an approximation for $\mathbf{G}(t)$ of order q (see [48], p. 174):

$$\mathbf{G}_q(t) = \sum_{j=1}^{q} \left[\mathbf{A}_j \cos(\nu_j t) + \mathbf{B}_j \sin(\nu_j t)\right], \tag{B.7}$$

where $\{\mathbf{A}_j, \mathbf{B}_j\}$, $j = 1, 2, \ldots, q$, is a zero-mean Gaussian random vector such that

$$\begin{aligned}
\mathrm{E}[A_{i,k}A_{j,l}] = \quad \mathrm{E}[B_{i,k}B_{j,l}] &= \delta_{ij} \int_{u_{i-1}}^{u_i} g_{kl}(\nu)\mathrm{d}\nu \approx \delta_{ij} g_{kl}(\nu_i)\Delta\nu_i, \\
\mathrm{E}[A_{i,k}B_{j,l}] = -\mathrm{E}[B_{i,k}A_{j,l}] &= \delta_{ij} \int_{u_{i-1}}^{u_i} h_{kl}(\nu)\mathrm{d}\nu \approx \delta_{ij} h_{kl}(\nu_i)\Delta\nu_i,
\end{aligned} \tag{B.8}$$

and $\delta_{ij} = 1$ if $i = j$ and zero otherwise, $i, j = 1, 2, \ldots, q$. By Eq. (B.7), $\mathbf{G}_q$ is a zero-mean, stationary Gaussian process for any $q \geq 1$. It can be shown that, as $q \to \infty$: (1) $\mathbf{G}_q$ approaches $\mathbf{G}$ in the mean-square sense, (2) the covariance functions of $\mathbf{G}_q$ approach the covariance functions of $\mathbf{G}$, and (3) $\mathbf{G}_q$ becomes a version (*i.e.*, has an identical finite dimensional distribution) of $\mathbf{G}$.

The generation of samples of $\mathbf{G}_q$ involves two steps. First, we generate samples of the zero-mean Gaussian vector, $\{\mathbf{A}_j, \mathbf{B}_j\}$, $j = 1, 2, \ldots, q$, with covariances defined by Eq. (B.8); methods from [48], Section 4.2, can be used for sample generation. Second, we apply Eq. (B.7), where we replace $\{\mathbf{A}_j, \mathbf{B}_j\}$ by the samples generated in step one.

Consider the case of sample generation of a zero-mean scalar (*i.e.*, $n = 1$) Gaussian process, denoted by G, with covariance function $c(\tau) = \mathrm{E}[G(t)G(t+\tau)]$. By Eqs. (B.7) and (B.8),

$$G_q(t) = \sum_{j=1}^{q} \left[A_j \cos(\nu_j t) + B_j \sin(\nu_j t)\right], \tag{B.9}$$

is an approximation for G, and $\{A_j, B_j\}$, $j = 1, 2, \ldots, q$, is a zero-mean Gaussian random vector such that, for $i, j = 1, 2, \ldots, q$,

$$\begin{aligned}
&\mathrm{E}[A_i B_j] = 0, \\
&\mathrm{E}[A_i A_j] = \mathrm{E}[B_i B_j] = \delta_{ij} \int_{u_{i-1}}^{u_i} g(\nu)\mathrm{d}\nu \approx \delta_{ij} g(\nu_i)\Delta\nu_i,
\end{aligned} \tag{B.10}$$

where, by Eq. (B.5),

$$g(\nu) = s_{11}(\nu) + s_{11}(-\nu) \tag{B.11}$$

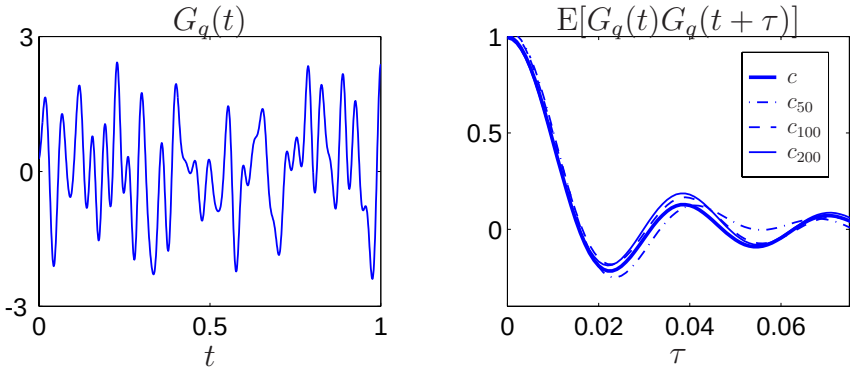

Figure B.1: One sample of G_q (left) and estimates of $c_q(\tau)$ (right).

is the one-sided PSD of G. Note that $s_{11}(\nu)$ is an even function of ν so that, by Eq. (B.5), $h_{11}(\nu) = 0$ and, by Eq. (B.8), $\mathrm{E}[A_i B_j] = 0$, $i, j = 1, 2, \ldots, q$.

For illustration, let

$$c(\tau) = \frac{\sin(\bar{\nu}\tau)}{\bar{\nu}\tau} \text{ and } g(\nu) = \frac{1}{\bar{\nu}} 1(\nu \leq \bar{\nu}). \tag{B.12}$$

Shown in Fig. B.1 is one sample of $G_q(t)$ for the case of $q = \bar{\nu} = 200$. Also shown are estimates of $c_q(\tau) = \mathrm{E}[G_q(t)G_q(t+\tau)]$, for the case of $q = 50$, $q = 100$, and $q = 200$ using 1,000 Monte Carlo samples. By Fig. B.1, we note that as q increases, $c_q(\tau)$ approaches $c(\tau)$ for all τ.

Appendix C

Second-moment properties of filtered Poisson processes

Recall the definition of the filtered Poisson process from Section 5.8.1

$$V(t;\lambda) = \begin{cases} 0 & N(t) = 0 \\ \sum_{k=1}^{N(t;\lambda)} W_k f(t - T_k) & N(t) > 0, \end{cases} \tag{C.1}$$

where $N(t;\lambda)$ is the counting process with jump times $\{T_k, k \geq 1\}$ and intensity $\lambda > 0$, the iid random variables $\{W_k, k \geq 1\}$ have zero mean and variance σ^2, and f is a known, deterministic shape function. Here, we consider the special case of homogenous $N(t)$, *i.e.*, λ is constant, and stationary V. Under these assumptions, the mean of V is given by (see [48], p. 82)

$$\begin{aligned} \mathrm{E}[V(t;\lambda)] &= \int_{-\infty}^{t} \mathrm{E}\left[W_1 f(t-s)\lambda\right] \mathrm{d}s \\ &= \lambda\, \mathrm{E}[W_1] \int_{-\infty}^{t} f(t-s)\mathrm{d}s \\ &= \lambda\, \mathrm{E}[W_1] \int_{0}^{\infty} f(u)\mathrm{d}u, \end{aligned} \tag{C.2}$$

where the last step follows with substitution $u = t - s$. Likewise, the variance is given by

$$\begin{aligned} \mathrm{Var}[V(t;\lambda)] &= \int_{-\infty}^{t} \mathrm{E}\left[\left(W_1 f(t-s)\right)^2 \lambda\right] \mathrm{d}s \\ &= \lambda\, \mathrm{E}\left[W_1^2\right] \int_{-\infty}^{t} \left[f(t-s)\right]^2 \mathrm{d}s \\ &= \lambda\, \sigma^2 \int_{0}^{\infty} f(u)^2 \mathrm{d}u. \end{aligned} \tag{C.3}$$

Assuming $\tau \geq 0$, the covariance of $V(t;\lambda)$ is

$$\begin{aligned}
\mathrm{E}[V(t;\lambda)V(t+\tau;\lambda)] &= \int_{-\infty}^{t} \mathrm{E}\left[W_1 f(t-s) W_1 f(t+\tau-s)\lambda\right] \mathrm{d}s \\
&= \lambda\ \mathrm{E}\left[W_1^2\right] \int_{-\infty}^{t} f(t-s) f(t+\tau-s) \mathrm{d}s \\
&= \lambda\ \sigma^2 \int_{0}^{\infty} f(u) f(u+\tau) \mathrm{d}u. \qquad \text{(C.4)}
\end{aligned}$$

The target mean, variance, and covariance of $V(t;\lambda)$ are zero, one, and $c(\tau)$ given by Eq. (5.13), respectively. The filtered Poisson process, $V(t;\lambda)$, will satisfy these requirements provided: (1) $\mathrm{E}[W_1] = 0$, (2) a $f(t) \in L_2$ can be found such that

$$\frac{1}{\bar{\gamma}} \int_{0}^{\infty} f(u) f(u+\tau) \mathrm{d}u = c(\tau), \qquad \text{(C.5)}$$

where

$$\bar{\gamma} = \int_{0}^{\infty} f(u)^2 \mathrm{d}u < \infty, \qquad \text{(C.6)}$$

and (3) $\lambda\sigma^2\bar{\gamma} = 1$.

The integral equation of Eq. (C.5) cannot be solved in closed-form. Let $\hat{f} \in L_2$ be an approximation to f where, for $u' > 0$, we define $\hat{f}(u) = 0$, $\forall u \notin [0, u']$. The following procedure is used to calculate $\hat{f}$:

1. choose $\tau' > 0$ such that $|c(\tau)| < \epsilon$ for all $\tau \geq \tau'$,
2. let $0 = \tau_0 < \tau_1 < \cdots < \tau_{n-1} < \tau_n = \tau'$ form a partition for $\tau \in [0, \tau']$,
3. let $0 = u_0 < u_1 < \cdots < u_{n-1} < u_n = u'$ form a partition for $u \in [0, u']$,
4. define
   (a) $t_i = \dfrac{\tau_i + \tau_{i-1}}{2}$, $i = 1, 2, \ldots, n$,
   (b) $\Delta u_i = u_i - u_{i-1}$, $i = 1, 2, \ldots, n$,
   (c) $\hat{f}_i = \hat{f}\left[\dfrac{u_i + u_{i-1}}{2}\right]$, $i = 1, 2, \ldots, n$,
5. the integral in Eq. (C.5), for $\tau = t_j$, is approximated as:

$$\int_{0}^{\infty} \hat{f}(u)\hat{f}(u+t_j)\mathrm{d}u \approx \sum_{i=1}^{n-j+1} \hat{f}_i \hat{f}_{i+j-1} \Delta u_i,\ j = 1, 2, \ldots, n, \qquad \text{(C.7)}$$

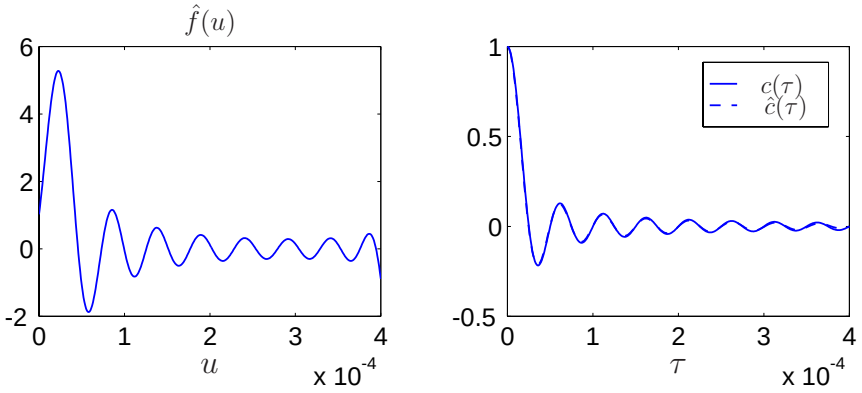

Figure C.1: Approximate solution, $\hat{f}$, (left) and corresponding covariance function for V (right).

6. Equation (C.6) is approximated as:

$$\int_0^\infty \hat{f}(u)^2 \mathrm{d}u \approx \sum_{i=1}^{n} \hat{f}_i^2 \Delta u_i, \tag{C.8}$$

7. define the residual at t_j, for $j = 1, 2, \ldots, n$, as:

$$r_j = \sum_{i=1}^{n-j+1} \hat{f}_i \hat{f}_{i+j-1} \Delta u_i - \sum_{i=1}^{n} \hat{f}_i^2 \Delta u_i c(t_j), \tag{C.9}$$

and

8. find $\hat{f}_i$, $i = 1, 2, \ldots, n$, such that $\sum_{j=1}^{n} r_j^2$ is a minimum.

Parameters $\tau' = u' = 4 \times 10^{-4}$, $\epsilon = 0.02$, and $n = 4000$ were used for calculations. The solution for $\hat{f}$ is shown on the left side of Fig. C.1; shown on the right are $\hat{c}(\tau)$, the solution to Eq. (C.5) with f replaced by $\hat{f}$, and the required covariance, $c(\tau)$. The errors are negligible for $\tau \in [0, \tau']$.

Appendix D

Algorithm to compute all paths through Ω

Recall the graph of Ω shown in Fig. 6.15. The collection of paths through Ω can be calculated by following the algorithm GetPaths, shown in Fig. D.1. Each path is consistent with the assumptions made in Section 6.2.1.2. The inputs to the algorithm are:

- the adjacency matrix ($\mathbf{c}$) for the graph,
- the node number for the treasure (treasure), and
- the node numbers for all entry nodes into Ω (entrynodes).

Output are:

- the number of paths (np), and
- a node list defining each path (paths).

The algorithm calculates each path in reverse order, *i.e.*, the paths begin at the treasure and end at one of the entry nodes; each path is then ordered from entry node to treasure at the end of the algorithm.

```
input  : treasure, c, entrynodes
output: np, paths

localnode = treasure
localpath = treasure
npr = # paths remaining in localpath
while npr > 0 do
    for j = 1 to npr do
        find all nodes one step away from localnode(j) using c
        append these nodes to localpath
    remove any repeated paths from localpath
    put paths that end at any entrynodes in paths and remove from
    localpath
    npr = # paths remaining in localpath
    localnode = last node in each remaining path
re-order each path in paths from one entrynodes to treasure
np = # paths in paths
```

Figure D.1: Algorithm GetPaths.

DISTRIBUTION:

1 Los Alamos National Laboratory
ESA-WR
MS:T080
Los Alamos, NM 87545
ATTN: Mark C. Anderson

1 Larry Bergman
Dept. of Aeronautical & Astronautical Engineering
University of Illinois
306 Talbot Lab
104 S. Wright St.
Urbana, IL 61801

1 Roger Ghanem
Dept. of Civil Engineering
Johns Hopkins University
201 Latrobe Hall
Baltimore, MD 21218

1 Mircea Grigoriu
Dept. of Civil Engineering
Cornell University
369 Hollister Hall
Ithaca, NY 14853

1 Los Alamos National Laboratory
ESA-WR
MS:T001
Los Alamos, NM 87545
ATTN: Francois Hemez

1 Erik Johnson
Dept. of Civil Engineering
University of Southern California
Los Angeles, CA 90089-2531

1 David Martinez
Strategic Systems Programs Liaison
SP-28E Nebraska Avenue Complex
287 Somers Court NW
Suite 10041
Washington DC 20393-5446

1 Katerina Papoulia
Dept. of Civil Engineering
Cornell University
363 Hollister Hall
Ithaca, NY 14853

1 Gennady Samorodnitsky
Dept. of Operations Research & Industrial Engineering
Cornell University
220 Rhodes Hall
Ithaca, NY 14853

1 MS 0370
M. Eldred, 9211

1 MS 0370
S. Mitchell, 9211

1 MS 0370
L. Swiler, 9211

1 MS 0370
T. Trucano, 9211

1 MS 0372
J. Jung, 9127

1 MS 0372
R. May, 9126

1 MS 0380
K. Alvin, 9142

1 MS 0380
G. Reese, 9142

1 MS 0384
T. Bickel, 9100

1 MS 0384
H. Morgan, 9140

1 MS 0451
J. Gauthier, 5533

1 MS 0557
T. Baca, 9125

1 MS 0557
T. Carne, 9124

1 MS 0653
B. Montano, 3520

1 MS 0824
J. Moya, 9130

1 MS 0825
W. Hermina, 9110

1 MS 0825
B. Blackwell, 9115

1 MS 0828
K. Dowding, 9133

1 MS 0828
T. Giunta, 9133

1 MS 0828
W. Oberkampf, 9133

1 MS 0828
T. Paez, 9133

1 MS 0828
M. Pilch, 9133

1 MS 0828
J. Red-Horse, 9133

1 MS 0828
V. Romero, 9133

1 MS 0828
A. Urbina, 9133

1 MS 0829
B. Rutherford, 12337

1 MS 0847
C. Dohrmann, 9124

10 MS 0847
R. Field, Jr., 9124

1 MS 0847
J. Redmond, 9124

1 MS 0847
D. Segalman, 9124

1 MS 0847
P. Wilson, 9120

1 MS 0847
S. Wojtkiewicz, 9124

1 MS 0893
J. Pott, 9123

1 MS 1004
M. McDonald, 15221

1 MS 1004
F. Oppel, 15221

1 MS 1138
S. Deland, 6222

1 MS 1351
R. Carlson, 5517

1 MS 9004
R. Stulen, 8100

3 MS 9018
Central Technical Files, 8945-1

2 MS 0899
Technical Library, 9616

1 MS 0612
Review & Approval Desk, 9612